ORIGINAL CONTRIBUTIONS

Airborne Particulate Matter and Mortality: Timescale Effects in Four US Cities

Francesca Dominici[1], Aidan McDermott[1], Scott L. Zeger[1], and Jonathan M. Samet[2]

[1] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.
[2] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

While time-series studies have consistently provided evidence for an effect of particulate air pollution on mortality, uncertainty remains as to the extent of the life-shortening implied by those associations. In this paper, the authors estimate the association between air pollution and mortality using different timescales of variation in the air pollution time series to gain further insight into this question. The authors' method is based on a Fourier decomposition of air pollution time series into a set of independent exposure variables, each representing a different timescale. The authors then use this set of variables as predictors in a Poisson regression model to estimate a separate relative rate of mortality for each exposure timescale. The method is applied to a database containing information on daily mortality, particulate air pollution, and weather in four US cities (Pittsburgh, Pennsylvania; Minneapolis, Minnesota; Seattle, Washington; and Chicago, Illinois) from the period 1987–1994. The authors found larger relative rates of mortality associated with particulate air pollution at longer timescale variations (14 days–2 months) than at shorter timescales (1–4 days). These analyses provide additional evidence that associations between particle indexes and mortality do not imply only an advance in the timing of death by a few days for frail individuals.

air pollution; Fourier analysis; hierarchical model; mortality; Poisson distribution; time factors; time series

Abbreviations: CI, confidence interval; PM_{10}, particulate matter with an aerodynamic diameter ≤ 10 μg/m³.

A number of studies over the last decade have shown an association between particle concentrations in outdoor air and daily mortality counts in urban locations (1–3). These associations have been estimated through the use of Poisson regression methods, and the findings have been reported as log relative rates of mortality associated with air pollution levels on recent days. These associations have been widely interpreted as reflecting the effect of air pollution on persons who have heightened susceptibility because of chronic heart or lung diseases (4).

Thus, the increased mortality associated with higher pollution levels may be restricted to very frail people whose life expectancy would have been short even without air pollution. This possibility is termed the "mortality displacement"

Correspondence to Dr. Francesca Dominici, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205 (e-mail: fdominic@jhsph.edu).

or "harvesting" hypothesis (5). If an effect is evident only at short timescales, pollution-related deaths are advanced by only a few days, and in fact, the days of life lost might arguably be of low quality for the frail individuals at risk of dying. Consequently, the public health relevance of the findings of the daily time-series studies has been questioned (6). The mortality displacement hypothesis received specific discussion in the 1996 Staff Paper on Particulate Matter prepared by the US Environmental Protection Agency because of its policy implications (4). The findings of two long-term prospective cohort studies of air pollution and mortality, the Harvard Six Cities Study (7) and the American Cancer Society's Cancer Prevention Study II (8), were considered to offer critical evidence counter to the mortality displacement hypothesis.

Several investigators have approached the problem of mortality displacement using analytical models for daily time-series data (9–12). If the association between air pollution and mortality does reflect the existence of a pool of frail individuals in the population, episodes of high pollution that lead to increased mortality might reduce the size of this pool, and days subsequent to high-pollution days would then be expected to show a reduced effect of air pollution. Therefore, the occurrence of this phenomenon can be investigated by assessing interaction between prior high-pollution days and the effects of subsequent pollution exposure on mortality counts; under the mortality displacement hypothesis, a negative interaction is predicted (13, 14).

Recently, Kelsall et al. (15) and Schwartz (11) developed related methods for analysis of daily time-series data, both offering approaches to estimating air pollution-mortality associations at varying timescales. More specifically, Kelsall et al.'s methodology gives a continuous smooth estimate of relative risk as a function of timescale (frequency domain log-linear regression). Zeger et al. (10) applied the frequency domain log-linear regression to previously analyzed data for Philadelphia, Pennsylvania, from 1973–1988. Schwartz (11, 12) used a filtering algorithm (16) to separate the time series of daily deaths, air pollution, and weather into long-wavelength components, midscale components, and residual, very short-term components and applied this method to data on Boston, Massachusetts, from 1979–1986 and Chicago, Illinois, from 1988–1993. Note that both of these methods (10, 11) analyze both pollution and mortality on the same timescales, i.e., shorter-term to longer-term. Both sets of analyses found effects on longer timescales.

In this paper, we extend the work by Zeger et al. (10) and Schwartz (11) in the methodological, substantive, and computational arenas. More specifically, we develop a timescale decomposition of a time series based on the discrete Fourier transform; we introduce a two-stage model for combining evidence across locations for estimation of pooled timescale-specific air pollution effects on mortality; and we provide the software for decomposing a time series into a set of desired timescale components. At the first stage of the model, we use Fourier series analyses (17, 18) to decompose the daily time series of the air pollution variable into distinct timescale components. This decomposition leads to a set of orthogonal predictors, each representing a specific timescale of variation in the exposure. We then use this set of predictors in Poisson

regression models to estimate a relative rate of mortality corresponding to each timescale exposure while controlling for other covariates such as temperature. A comparison between our approach and the frequency domain log-linear regression analysis is provided below in the section "Sensitivity analysis and model comparison."

The method is applied to concentrations of particulate matter, based on measurements of particles with an aerodynamic diameter less than or equal to 10 $\mu g/m^3$ (PM_{10}) and daily mortality counts from four US cities—Pittsburgh, Pennsylvania; Minneapolis, Minnesota; Chicago, Illinois; and Seattle, Washington. These were four cities with daily PM_{10} measurements that were among the 90 largest US cities used in the National Morbidity, Mortality, and Air Pollution Study (19, 20). The analyses are restricted to these cities because they are the only US locations with daily air pollution concentrations available in this database for this time interval, while in most other locations, PM_{10} levels were measured only every 6 days as required by the Environmental Protection Agency. Our approach is not suitable for every-sixth-day PM_{10} data, for two reasons: 1) no information is available from the data for estimation of the short-term effects of air pollution on mortality and 2) because of the "aliasing" phenomenon, the effects of air pollution at the longer timescales are distorted. In our context, the aliasing phenomenon occurs when the sampling interval is larger than 1 day, so that variations in the daily time series at the shortest timescales produce an apparent effect at the longer timescales.
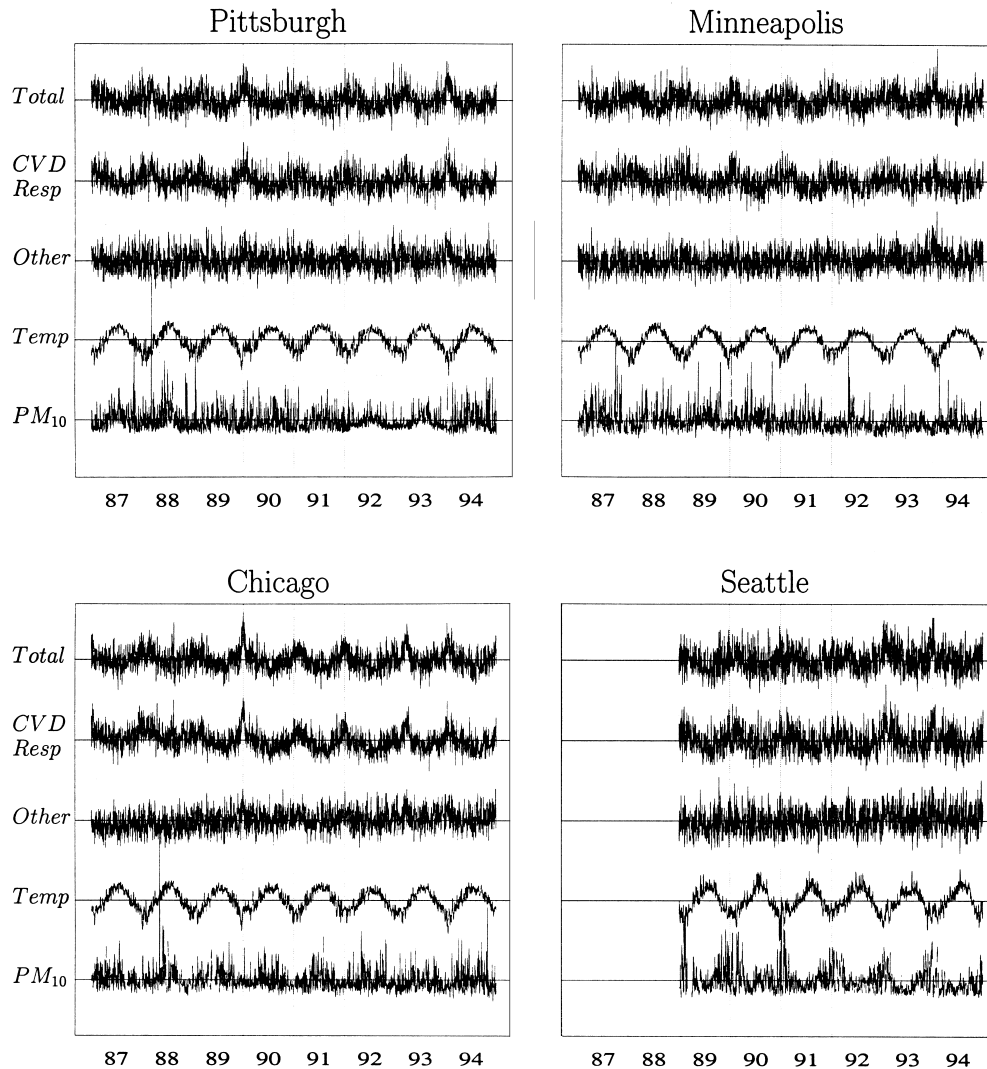
MATERIALS AND METHODS

Data

We used daily time series of mortality, weather, and air pollution data for Pittsburgh, Minneapolis, Chicago, and Seattle for the period 1987–1994 (see figure 1). Daily mortality counts were obtained from the National Center for Health Statistics and were grouped by age (<65, 65–75, and >75 years) and by cause of death according to the *International Classification of Diseases*, Ninth Revision (cardiovascular-respiratory mortality (cardiac conditions, codes 390–448; respiratory conditions, codes 490–496; influenza, code 487; and pneumonia, codes 480–486, 507) and mortality due to other remaining diseases). Accidental deaths were excluded. Hourly temperature and dew point data were available from the National Climatic Data Center, assembled in a compact disk database from EarthInfo, Inc. (21). The air pollution data were obtained from the Aerometric Information Retrieval Service (22) database maintained by the Environmental Protection Agency. For the pollutants measured on an hourly basis, we calculated the 24-hour average. A more detailed description of the database has been published elsewhere (19, 20).

Methods

Below we describe our statistical approach to estimation of the association between air pollution and mortality using different timescales. We let X_t^c be the air pollution time

FIGURE 1.   Daily time series of mortality (total, cardiovascular disease (CVD) and respiratory (Resp), and other causes (Other)), temperature (Temp), and levels of particulate matter with an aerodynamic diameter less than 10 µg/m³ (PM$_{10}$) for Pittsburgh, Pennsylvania, Minneapolis, Minnesota, and Chicago, Illinois, during the period 1987–1994. For Seattle, Washington, data for the period 1989–1994 were used.
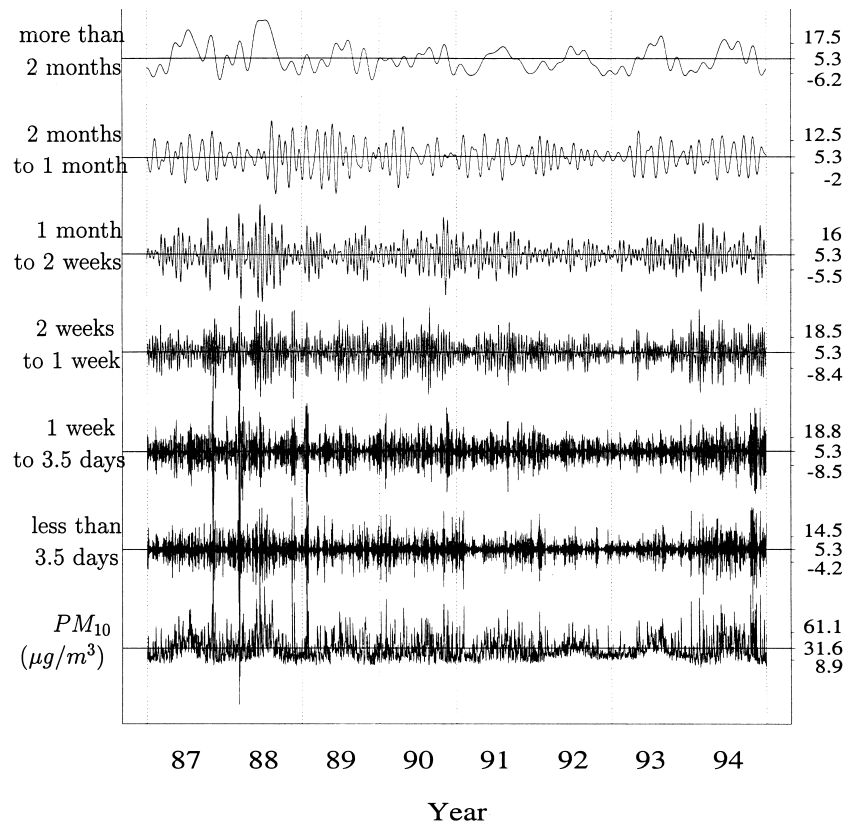
series and Y_t^c be the mortality time series in location c. We first decompose the air pollution series X_t^c into distinct component series X_{kt}^c, one for each distinct timescale k, and then we calculate the association between Y_t^c, without decomposition, and each of the timescale components X_{kt}^c. The decomposition is obtained by applying the discrete Fourier transform to the X_t^c series (17, 18). Specifically, we assume

$$Y_t^c \big| \mu_t^c \sim \text{Poisson}(\mu_t^c); \ \text{var}(Y_t^c) = \varphi^c \mu_t^c$$

and

$$\log \mu_t^c = \Sigma_k X_{kt}^c \beta_k^c + S(\text{time}, \ 7/\text{year}) + \text{confounders}, \quad (1)$$

where φ^c denotes the overdispersion parameter and the β_k^c's, the parameters of interest, denote the log relative rate of daily mortality for each 10-unit increase in the air pollution level in location c on a timescale k. Our modeling approach replaces the term $X_t^c \beta^c$, where X_t^c is the air pollution time series and β^c is the city-specific log relative rate of mortality, with the sum $\Sigma_k X_{kt}^c \beta_k^c$, where $X_t = \Sigma_k X_{kt}$, and the $X_{1t}, \ldots, X_{kt}, \ldots, X_{Kt}$ is a set of orthogonal predictors. This model estimates relative rates of mortality at different timescales and characterizes the timescale variation in the air pollution time series that contributes to the estimate of the overall effect β^c. Here we expect that under a short-term mortality displacement scenario, mortality would be mainly associated with a short-term effect of air pollution.

FIGURE 2. Decomposition into a six-component series of data on particulate matter with an aerodynamic diameter ≤10 µg/m³ (PM$_{10}$) for Pittsburgh, Pennsylvania, from the period 1987–1994. The bottom row shows the observed PM$_{10}$ values with a horizontal line plotted at the mean value (31.6 µg/m³) and tick marks at the 10th (8.9 µg/m³) and 90th (61.1 µg/m³) percentiles. Time series 1 to 6 (top to bottom) are the timescale decompositions of the observed data series. A horizontal line is plotted at the mean value, which in all cases is one sixth of the original mean (5.3 µg/m³), and tick marks show the 10th and 90th percentiles (see vertical axis at right). The decomposition is obtained by applying the function *decompose*( ) with breaks equal to 1, 48, 96, 208, 416, 834, and 2,922 days. The internal breaks (48, 96, 208, 416, and 834) were obtained by dividing the length of the data (e.g., T = 2,922 days) by timescales of 60, 30, 14, 7, and 3.5 days. One should always select the minimum and maximum breaks to be 1 and T in order to capture the lowest and highest Fourier frequencies.

To protect the pollution relative rates from confounding by longer-term trends and seasonality, we also remove the variation in the time series at timescales approximately longer than 2 months by including a smooth function of time with 7 degrees of freedom per year. A sensitivity analysis with respect to selection of the number of degrees of freedom in the smooth function of time is discussed below. Smooth functions of temperature and dew point temperature are used to control for potential confounding by temperature and humidity. The rationale for and details on the selected smooth functions are provided by Samet et al. (23–25), Kelsall et al. (26), and Dominici et al. (27).
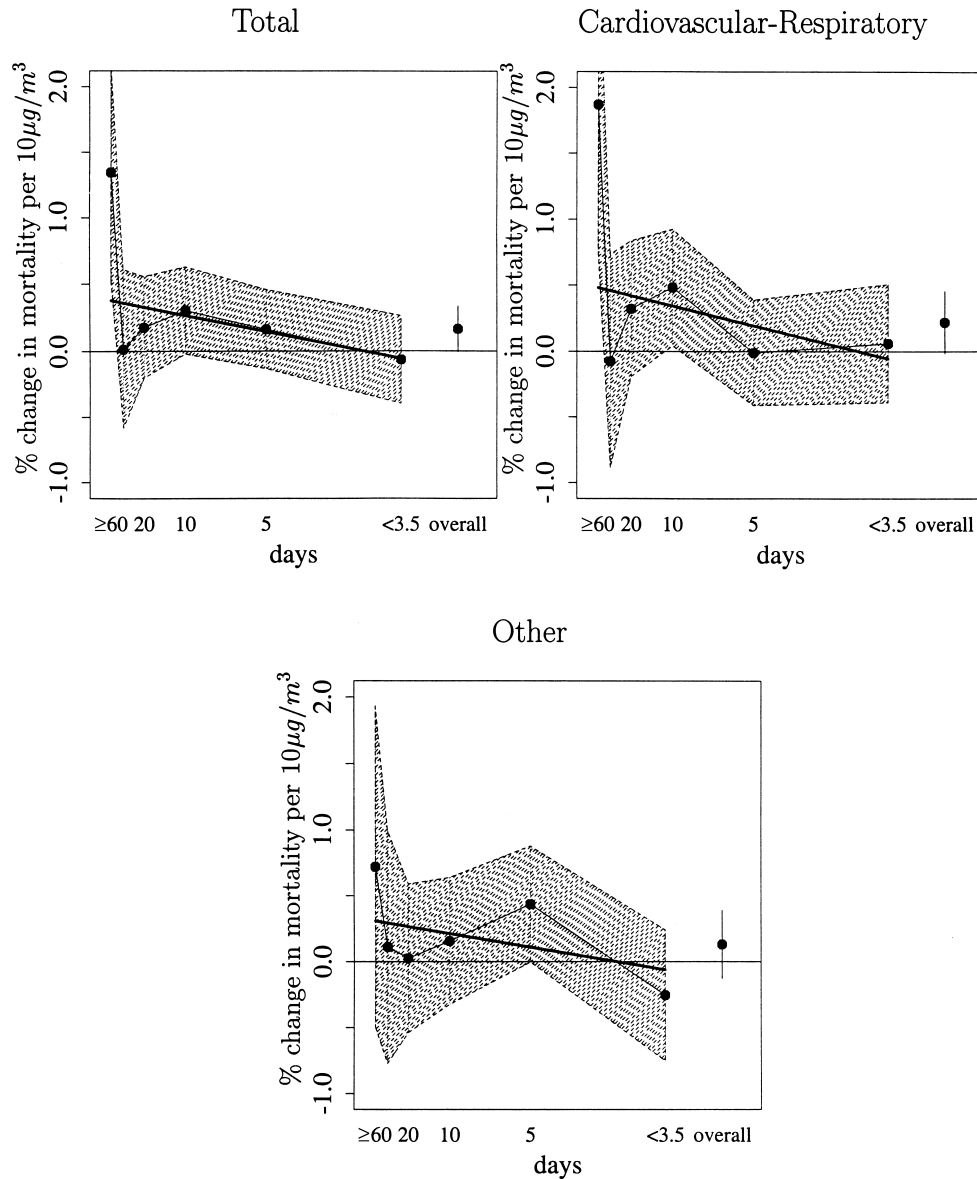
Figure 2 illustrates the decomposition of X_t into six timescales. From the top of the panel to the bottom are displayed time series ranging from series that comprise only the more smooth fluctuations (low frequency components) to time series that comprise only the less smooth variations (high frequency components). The actual value of X_t (the last time series at the bottom) on day t is obtained by summing the values of the six component series on each day. Details on the Fourier series decomposition and the URL address for downloading the software for its implementation are provided in the Appendix. Using the decomposed time series, we can estimate timescale coefficients ($\hat{\beta}_k^c$) denoting the relative change in mortality per 10-µg/m³ increase in the corresponding timescale components of PM$_{10}$ (X_{kt}^c). We estimate a vector of regression coefficients $\hat{\beta}^c = [\hat{\beta}_1^c, \dots, \hat{\beta}_K^c]$ and their covariance matrix V^c. Although one might want to allow a latency time for the effect of pollution, in equation 1 we regress Y_t^c on X_{kt}^c rather than use lagged pollution series $X_{kt-l_k}^c$ for some lag $l_k > 0$. We investigate whether a lagged predictor is needed in the sensitivity analysis.

RESULTS

Estimation

To estimate the timescale effects $\hat{\beta}^c$ and its covariance matrix V^c within each location c, we fit the Poisson regression model (equation 1) with smooth functions of time, temperature, and dew point modeled as natural cubic splines and using the function glm( ) in S-Plus (MathSoft, Inc., Cambridge, Massachusetts). To combine the coefficients

FIGURE 3. Pooled estimates of log relative rates of mortality at different timescales for total mortality (top left), cardiovascular-respiratory mortality (top right), and mortality due to other causes (bottom) according to levels of particulate matter with an aerodynamic diameter ≤10 µg/m³ in four US cities, 1987–1994. The estimates are plotted at ≥60, 40 (not shown on x-axis), 20, 10, 5, and <3.5 days, denoting approximate midpoints of the intervals ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days, respectively. The shaded regions represent ±2 standard errors of the estimates. Bars, 95% confidence interval.

across cities, we use a fixed-effect model with weights $W^c = (V^c)^{-1}$ and an estimator of the form

$$\bar{\beta} = (\Sigma_c W^c)^{-1} \sum_c W^c \hat{\beta}^c$$

with variance

$$V(\bar{\beta}) = (\Sigma_c W^c)^{-1}.$$

An alternative approach would be to use as weights $W^c = (D + V^c)^{-1}$, where D is a diagonal between-city covariance matrix with diagonal element τ^2. Because of the limited number of cities in the present analysis, we cannot estimate τ^2 reliably and have assumed $\tau = 0$. A sensitivity analysis of our results with respect to different values of τ^2 obtained from hierarchical analyses of data from 20 cities (27) and 88 cities (28) is discussed below.

We estimated city-specific and pooled log relative rates of mortality for the following six timescale variations of PM₁₀: ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days. Figure 3 shows the pooled estimates of the log relative rates of mortality at different timescales for total

TABLE 1. Timescale and city-specific estimates of log relative rates of mortality (total, cardiovascular and respiratory, and other causes) associated with a 10-unit increase in particulate matter with an aerodynamic diameter less than or equal to 10 µg/m³ in four US cities, 1987–1994

Type of mortality and timescale* (days)	Pittsburgh, Pennsylvania		Minneapolis, Minnesota		Chicago, Illinois		Seattle, Washington		Pooled data	
	Log RR†	95% CI†	Log RR	95% CI	Log RR	95% CI	Log RR	95% CI	Log RR	95% CI
Total mortality										
≥60	1.34	0.18, 2.86	0.51	−1.90, 2.93	1.57	0.37, 2.77	1.37	−0.98, 3.72	1.35	0.52, 2.17
40	0.95	−0.23, 2.13	−1.11	−2.97, 0.75	−0.14	−0.97, 0.69	−0.33	−1.98, 1.32	0.01	−0.58, 0.61
20	−0.57	−1.30, 0.15	1.21	−0.10, 2.51	0.36	−0.15, 0.87	0.38	−0.92, 1.69	0.18	−0.20, 0.56
10	0.19	−0.45, 0.83	0.06	−1.03, 1.15	0.42	−0.01, 0.84	0.11	−1.19, 1.40	0.30	−0.02, 0.63
5	0.27	−0.27, 0.81	0.92	−0.12, 1.95	−0.01	−0.40, 0.37	0.43	−1.15, 2.00	0.16	−0.13, 0.46
<3.5	0.09	−0.57, 0.75	−0.67	−1.86, 0.52	−0.06	−0.47, 0.36	0.02	−2.30, 2.33	−0.07	−0.40, 0.27
Overall	0.15	−0.19, 0.49	0.24	−0.37, 0.84	0.16	−0.07, 0.38	0.26	−0.45, 0.98	0.17	−0.01, 0.34
Cardiovascular and respiratory mortality										
≥60	1.19	−0.86, 3.24	0.99	−2.36, 4.35	2.54	0.90, 4.17	1.86	−1.44, 5.15	1.87	0.75, 2.99
40	1.00	−0.59, 2.59	−1.80	−4.38, 0.78	0.35	−0.78, 1.48	−2.67	−4.95, −0.39	−0.08	−0.89, 0.74
20	−0.33	−1.31, 0.64	1.00	−0.80, 2.79	0.45	−0.24, 1.14	1.04	−0.76, 2.85	0.32	−0.19, 0.84
10	0.44	−0.42, 1.31	−0.12	−1.62, 1.38	0.63	0.05, 1.20	0.24	−1.56, 2.03	0.48	0.04, 0.92
5	0.22	−0.50, 0.95	0.51	−0.92, 1.94	−0.24	−0.77, 0.28	0.84	−1.34, 3.01	−0.01	−0.41, 0.39
<3.5	0.23	−0.65, 1.11	−0.48	−2.13, 1.16	0.05	−0.51, 0.61	0.22	−2.99, 3.44	0.06	−0.39, 0.51
Overall	0.25	−0.20, 0.70	0.07	−0.77, 0.90	0.23	−0.08, 0.53	0.22	−0.76, 1.21	0.22	−0.02, 0.46
Mortality due to other causes										
≥60	1.45	−0.85, 3.74	−0.83	−4.36, 2.69	0.75	−1.02, 2.52	0.56	−2.83, 3.96	0.72	−0.50, 1.94
40	0.85	−0.92, 2.61	0.02	−2.70, 2.73	−0.74	−1.98, 0.50	2.04	−0.38, 4.47	0.11	−0.77, 0.99
20	−0.93	−2.01, 0.16	1.26	−0.65, 3.17	0.31	−0.44, 1.07	−0.23	−2.13, 1.68	0.03	−0.54, 0.59
10	−0.10	−1.06, 0.86	0.33	−1.26, 1.93	0.22	−0.41, 0.84	0.30	−1.59, 2.19	0.16	−0.33, 0.64
5	0.40	−0.42, 1.21	1.39	−0.13, 2.90	0.32	−0.25, 0.90	0.28	−2.02, 2.58	0.44	0.00, 0.88
<3.5	−0.14	−1.14, 0.86	−1.03	−2.77, 0.72	−0.20	−0.81, 0.42	−0.45	−3.82, 2.91	−0.26	−0.75, 0.24
Overall	0.02	−0.49, 0.53	0.38	−0.50, 1.27	0.11	−0.23, 0.45	0.38	−0.66, 1.43	0.13	−0.13, 0.39

* The numbers given for the timescales denote the midpoints of the following intervals: ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days.

† RR, relative rate; CI, confidence interval.

mortality, cardiovascular and respiratory mortality, and mortality due to other causes. At the far right are the plotted estimates of the log relative rate of mortality obtained using the nondecomposed time series X_t. For all causes and for cause-specific mortality, we found that estimates of the association between air pollution and mortality obtained using the smoother variations in the time series (10 days to 1 month) are somewhat larger than those obtained using the less smooth variations (1–3 days). The largest effects occurred at timescales greater than 2 months for total mortality (1.35 percent per 10 µg/m³; 95 percent confidence interval (CI): 0.52, 2.17), cardiovascular and respiratory mortality (1.87 percent per 10 mg/m³; 95 percent CI: 0.75, 2.99), and other-cause mortality (0.72 percent per 10 µg/m³; 95 percent CI: –0.55, 1.95). To test the hypothesis that estimated effects at the longer timescales are larger than those at the shortest timescales, we linearly regress the

pooled $\hat{\beta}_k$ on the timescales and calculated weighted least squares estimates. The solid lines in figure 3 represent the fitted linear regressions. In all cases, the estimated slopes are negative, with t statistics close to the significance level.

Table 1 shows the estimates of the timescale effect for each city. At the city-specific level, we found a similar pattern, with larger effects at the longer timescales and smaller effects at the shorter timescales, with few exceptions. Consistent with previous studies (19, 20), the log relative rates, obtained by averaging across cities and across timescales, are highest for cardiovascular and respiratory mortality (0.22 percent per 10 µg/m³; 95 percent CI: –0.02, 0.46) and lowest for other-cause mortality (0.13 percent per 10 µg/m³; 95 percent CI: –0.13, 0.39). These findings suggest that the estimates of association between air pollution and mortality at the greater timescales contribute substantially to the estimation of the overall log relative rate β^c.

TABLE 2.   Analysis of the sensitivity of estimated log relative rates of mortality due to air pollution to the choice of lag for exposure series at timescales shorter than 1 month in four US cities, 1987–1994*

City	14–29 days	7–13 days	3.5–6 days	<3.5 days
Pittsburgh, Pennsylvania	1	2	0	6
Minneapolis, Minnesota	0	5	2	3
Chicago, Illinois	0	0	6	5
Seattle, Washington	0	3	6	4

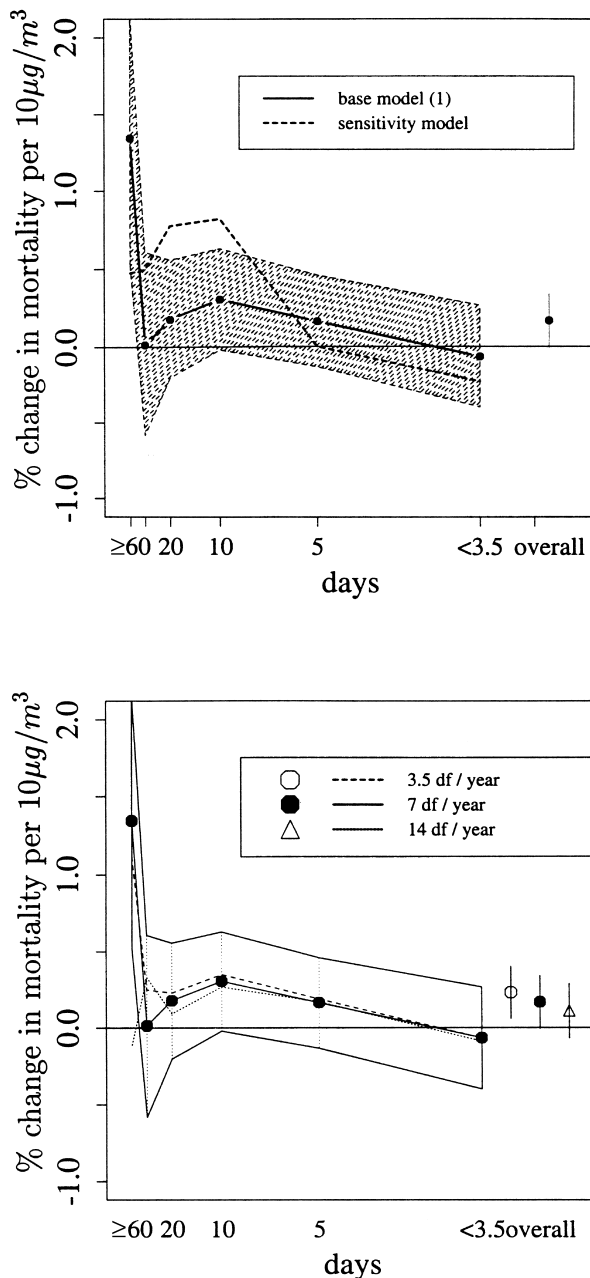* The table summarizes, for each city and each timescale, the lag associated with the largest *t* statistic.

Sensitivity analysis and model comparison

Below we investigate the sensitivity of our results with respect to: 1) lag choice in the air pollution time series; 2) adjustment for long-term trends and seasonality; and 3) the degree of heterogeneity of the true relative rates among cities. We also apply our timescale approach to the Philadelphia database for 1973–1988 that was previously analyzed by Zeger et al. (10), to compare methods used here with the frequency domain log-linear regression estimates previously published.

We first test the sensitivity of the log relative rate estimates to the choice of lag for component exposure series at timescales shorter than 1 month. We assume that the lag l_k is 0 for timescales greater than 1 month, since lags of 4 days will have little effect on the results for large timescales. We fit several different lags for each component exposure series X_{kt-l_k} and choose the best lag l_k (the one with the largest *t* statistic) rather than assume that l_k is 0. The optimal lags were obtained by including all timescale components in the models; they are summarized in table 2. Results for total mortality with an optimal lag compared with the original model with a zero lag are shown in the upper half of figure 4. Although the estimates differ at particular timescales, the overall shape of the curves remains similar and remains inconsistent with the short-term mortality displacement hypothesis.

Our model controls for long-term trends in mortality by including a natural cubic spline of time with 7 degrees of freedom per year. To assess the sensitivity of the results to the choice of smoothing parameter, we repeat the analysis using 3.5 and 14 degrees of freedom per year. The lower half of figure 4 shows the resulting plots. As expected, when less flexible curves are used (fewer degrees of freedom), there is less control for trend and seasonality, and therefore the effects at the longer timescales tend to be higher.

Our strategy for investigating the impact of the assumption of homogeneity ($\tau = 0$) of the pollution effects on our results is based on inspecting the pooled timescale estimates for total mortality under four alternative values for τ. These were extracted from Bayesian hierarchical analyses of 20 cities (27) and the 88 largest cities in the United States (28). The posterior mean values of τ and the corresponding prior distributions are summarized in table 3.

FIGURE 4. Top: Sensitivity analysis showing stepwise pooled estimates for the total mortality and PM$_{10}$ (particulate matter with an aerodynamic diameter ≤10 μg/m^3) data series (four US cities, 1987–1994) and model 1 (see equation 1 in text). Bottom: Sensitivity analysis showing pooled estimates for the total mortality and PM$_{10}$ data series using different numbers of degrees of freedom (df) for time but the same smooth function for temperature and humidity. The estimates are plotted at ≥60, 40 (not shown on *x*-axis), 20, 10, 5, and <3.5 days, denoting approximate midpoints of the intervals ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days, respectively. Bars, 95% confidence interval.

Results are shown in table 4. The results were all obtained by using smooth functions (natural cubic splines) of time with 7 degrees of freedom per year and smooth functions (natural cubic splines) of temperature and dew point as

TABLE 3. Sensitivity analysis of pooled estimates of the log relative rate of mortality due to air pollution with respect to the amount of heterogeneity (τ)*

$E[\tau|\text{data}]$	Prior distribution	Published reference
0.15	$\tau^2 \sim N\dagger(0, 305)I_{\tau^2 > 0}$	20-city analysis (discussion and rejoinder of Dominici et al. (27))
0.38	$\tau^2 \sim N(0, 1)I_{\tau^2 > 0}$	88-city analysis (Dominici et al. (28))
0.49	$\tau^2 \sim IG\ddagger(3, 1)$	88-city analysis (Dominici et al. (28))
0.76	$\tau^2 \sim IG(3, 6)$	20-city analysis (Dominici et al. (27))

* The table summarizes posterior mean values of τ used to calculate the pooled effect.
† N denotes the normal distribution.
‡ IG denotes the inverse gamma distribution.

confounders, as in model 1 (equation 1). Although the confidence intervals widen considerably, the pooled estimates are not very sensitive to the different values of τ. Even in the presence of substantial heterogeneity of the relative rates of mortality across cities, we still found that the pooled estimates at the longer timescales are larger than the pooled estimates at the shorter timescales, but with larger standard errors.

We now apply timescale and frequency domain log-linear regression analyses (10, 15) to the Philadelphia data set, and we estimate relative rates of mortality for exposure to air pollution at different timescales by using the Poisson regression model defined in equation 1. Figure 5 shows the frequency domain estimate of the mortality relative rate associated with air pollution as a function of Fourier frequency. Similar to the timescale result, the horizontal axes denote the Fourier frequencies (lower *x*-axis) and the timescale in days (upper *x*-axis) at which the association is measured. The solid curve and dotted curves denote the estimated relative rates ±2 estimated standard errors at each frequency. The timescale estimates (points connected by line segments) are plotted on top of the frequency domain results (continuous curve). Timescale estimates and frequency domain results are similar, and consistently with our results for the four cities, relative rate estimates at longer timescales are larger than relative rate estimates at short timescales.

DISCUSSION

This paper provides additional evidence that the association between particle indexes and mortality is greater at longer timescales (10 days to 2 months) than at timescales of a few days. This suggests that the association of air pollution with daily mortality counts does not reflect short-term mortality displacement alone. More specifically, our results are inconsistent with the "harvesting only" hypothesis, which contends that the air pollution-mortality association is caused entirely by frail persons' dying a few days earlier than they would have absent pollution. Under that hypothesis, we would anticipate little or no association at longer timescales. In fact, we observe the strongest associations there.

The larger relative rates at longer timescales may partly reflect a greater biologic impact on chronic exposures than on acute exposures. In fact, estimated relative risks from the Harvard Six Cities (7) and American Cancer Society (8) cohort studies, which address chronic exposures, are larger than estimates from times-series models (28), which are constrained to estimate the effects of shorter-timescale exposures.

The estimated relative rate of total mortality for the longest timescale (2 months) was 1.35 percent per 10-unit increase in PM_{10} for the four cities considered. While 1.35 percent is approximately 8 times larger than the overall pooled esti-
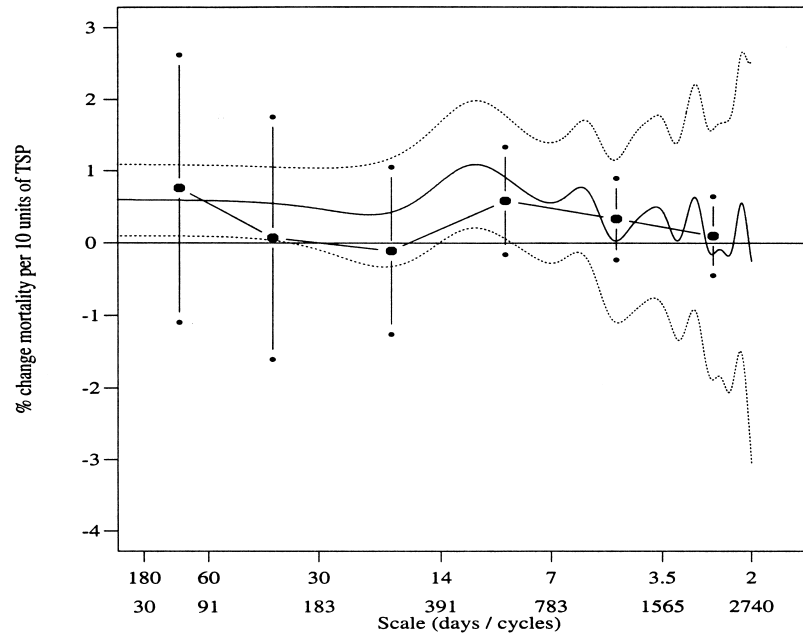
TABLE 4. Pooled estimates of the log relative rate of mortality due to air pollution under different values of the heterogeneity parameter τ

Timescale* (days)	$\tau = 0$		$\tau = 0.15$		$\tau = 0.38$		$\tau = 0.49$		$\tau = 0.76$	
	Log RR†	95% CI†	Log RR	95% CI	Log RR	95% CI	Log RR	95% CI	Log RR	95% CI
≥60	1.35	0.52, 2.17	1.34	0.50, 2.18	1.32	0.39, 2.25	1.31	0.32, 2.30	1.29	0.12, 2.45
40	0.01	−0.58, 0.61	0.01	−0.61, 0.63	0.00	−0.74, 0.73	−0.02	−0.83, 0.79	−0.05	−1.06, 0.96
20	0.18	−0.20, 0.56	0.18	−0.24, 0.60	0.21	−0.36, 0.79	0.24	−0.43, 0.90	0.28	−0.61, 1.18
10	0.30	−0.02, 0.63	0.28	−0.09, 0.66	0.24	−0.31, 0.79	0.23	−0.41, 0.87	0.21	−0.66, 1.09
5	0.16	−0.13, 0.46	0.20	−0.15, 0.55	0.29	−0.25, 0.83	0.32	−0.32, 0.95	0.36	−0.52, 1.23
<3.5	−0.07	−0.40, 0.34	−0.07	−0.46, 0.32	−0.11	−0.69, 0.47	−0.12	−0.80, 0.56	−0.14	−1.07, 0.78
Overall	0.17	−0.01, 0.34	0.18	−0.07, 0.43	0.19	−0.26, 0.64	0.19	−0.35, 0.74	0.20	−0.60, 1.00

* The numbers given for the timescale denote the midpoints of the following intervals: ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days.
† RR, relative rate; CI, confidence interval.

FIGURE 5. Change in mortality according to level of total suspended particulates (TSP) (particulate matter with an aerodynamic diameter ≤10 μg/m³) in Philadelphia, Pennsylvania, 1973–1988. The figure depicts a comparison between frequency domain estimates (continuous curve) and timescale estimates (points connected by line segments), showing the log relative rates of total mortality by frequency and frequency grouping. The dotted lines show ±2 standard errors for the frequency domain estimates, and the bars represent ±2 standard errors for the timescale estimates.

mate of 0.17 percent, it is still an order of magnitude smaller than the estimated relative risks from the cohort studies (7, 8). Thus, the time-series relative rates, even when restricted to longer-term exposures, are much smaller than those from the major cohort studies. This difference might indicate that the most harmful exposures occur over much larger timescales than can be studied with time-series methods. However, relative rate estimates at the longer timescales should be interpreted with caution because of the confounding effects of seasonality and trend.

Our results are consistent with findings from previous reports for Philadelphia (10), Boston (11), and Chicago (12) that have used harvesting-resistant estimators. These methods are based on a conceptually straightforward stratification of the air pollution time series into different frequency bands, allowing assessment of associations on timescales with differing implications.

Our approach and the approaches proposed by Zeger et al. (10) and Schwartz (11, 12) address related but different questions. Zeger et al. (10) and Schwartz (11, 12) decompose both the air pollution time series and the mortality time series into different timescales of variation (X_{kt} and Y_{kt}) and then aim to identify the timescale component that leads to the strongest association between time-averaged air pollution and time-averaged mortality. The timescale analysis proposed in this paper decomposes only the air pollution time series into different timescales (X_{kt}) and then characterizes the timescale variation of the effect of exposure on daily mortality. For linear models, these two approaches will provide the same

results. In Poisson regression, with small effects such as those that occur with air pollution variables, the differences between results from the two approaches will probably be small. Our approach, however, is applicable over the range of Poisson or other generalized linear model applications.

The timescale decomposition shown in figure 2 could have been performed using wavelet methods. Wavelets are a natural extension of Fourier analysis; however, in wavelet analysis, the window or "scale" with which we look at the information stream is selected automatically. In our context, this automatic selection of the timescales is not particularly desirable. One of the advantages of using wavelets is that functions with discontinuities and functions with sharp spikes can be represented using substantially fewer wavelet basis elements than sine-cosine basis elements. Because a common characteristic of time series of mortality, air pollution, and weather data is their periodicity without large discontinuities, Fourier analysis is adequate for our purpose.

The mortality displacement problem that motivated the development of this method is not unique to air pollution; it has also been discussed in relation to heat waves and influenza. The statistical approach proposed in this paper is suitable for these or other epidemiologic analyses with the focus of differentiating short-term effects from long-term effects of a time-varying exposure on a health outcome. The set of timescales selected should match hypotheses concerning relations between exposure time and response. We also provide an alternative strategy with which to control for

temporal confounding, since it is likely that confounding may vary with the timescale.

The timescale estimates from model 1 lead to specific patterns for the coefficients of a distributed lag model (9, 29, 30). A large effect at timescale k corresponds to an increased number of deaths for $k/2$ days after an air pollution episode, followed by a rebound below the baseline level for another $k/2$ days, owing to the depletion of the pool of susceptible people.

Unlike the distributed lag model, our approach is symmetric in time; that is, we use a symmetric time window $(t - l_k, t + l_k)$ to estimate X_{kt}. The temporal symmetry of our approach does not complicate our inferences, for two reasons. First, and most importantly, it is not plausible that mortality causes air pollution; it is only reasonable to consider the possibility that air pollution causes mortality. Second, we use a symmetric time window simply to better estimate the smooth variations of air pollution X_{kt}.

Other key methodological issues in time-series studies of air pollution and mortality are the nonlinearity in the dose-response curves, the effect of copollutants, and the effect of measurement error. These issues are discussed elsewhere and remain a topic of investigation (25, 28, 31, 32). In the context of mismeasurement of exposure, it is expected that the relative rate of mortality corresponding to the short time scales might be more attenuated by the measurement error than the relative rate of mortality corresponding to longer timescales. This is because more of the short timescale signal is actually error, whereas the longer timescale measure has effectively been smoothed so that measurement error is less of a contributor and hence less a source of bias. However, measurement error will not reverse the sign of an estimated coefficient or reverse the shape of the curve in figure 3. Therefore, even in the presence of measurement error, our results still do not support the "harvesting hypothesis" that the association between particle concentrations and mortality is entirely due to mortality among very frail persons who lose a few days of life.

REFERENCES

1. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. Health effects of outdoor air pollution. [Part 1]. Am J Respir Crit Care Med 1996;153:3–50.
2. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. Health effects of outdoor air pollution. Part 2. Am J Respir Crit Care Med 1996; 153:477–98.
3. Pope CA III, Dockery DW, Schwartz J. Review of epidemiological evidence of health effects of particulate air pollution. Inhal Toxicol 1995;7:1–18.
4. Environmental Protection Agency, Office of Air Quality Planning and Standards. Review of the National Ambient Air Quality Standards for Particulate Matter: policy assessment of scientific and technical information. OAQPS Staff Paper. Research Triangle Park, NC: Environmental Protection Agency, 1996. (Publication no. EPA-452\R-96-013).
5. Schimmel H, Murawski TJ. Proceedings: the relation of air pollution to mortality. J Occup Med 1976;18:316–33.
6. Lipfert FW, Wyzga RE. Air pollution and mortality: issues and uncertainties. J Air Waste Manage Assoc 1995;45:949–66.
7. Dockery DW, Pope CA III, Xu X, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med 1993;329:1753–9.
8. Pope CA III, Thun MJ, Namboordiri MM, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. Am J Respir Crit Care Med 1995;151:669–74.
9. Zanobetti A, Schwartz J, Samoli E, et al. The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. Epidemiology 2002;13:87–93.
10. Zeger SL, Dominici F, Samet J. Harvesting-resistant estimates of air pollution effects on mortality. Epidemiology 1999;10: 171–5.
11. Schwartz J. Harvesting and long term exposure effects in the relationship between air pollution and mortality. Am J Epidemiol 2000;151:440–8.
12. Schwartz J. Is there harvesting in the association of airborne particles with daily deaths and hospital admissions? Epidemiology 2001;12:55–61.
13. Spix C, Heinrich J, Dockery D, et al. Air pollution and daily mortality in Erfurt, East Germany, 1980–1989. Environ Health Perspect 1993;101:518–26.
14. Smith RL, Davis JM, Speckman P. Human health effects of environmental pollution in the atmosphere. In: Barnett V, Stein A, Turkman F, eds. Statistics in the environment 4: statistical aspects of health and the environment. Chichester, United Kingdom: John Wiley and Sons Ltd, 1999:91–115.
15. Kelsall J, Zeger S, Samet J. Frequency domain log-linear models: air pollution and mortality. Appl Stat 1999;48:331–44.
16. Cleveland WS, Develin SJ. Robust locally-weighted regression and smoothing scatterplots. J Am Stat Assoc 1988;74:829–36.
17. Bloomfield P. Fourier analysis of time series: an introduction. New York, NY: John Wiley and Sons, Inc, 1976.
18. Priestley MB. Spectral analysis and time series. New York, NY: Academic Press, Inc, 1981.
19. Samet JM, Zeger SL, Dominici F, et al. The National Morbidity, Mortality, and Air Pollution Study. Part II: morbidity and

mortality from air pollution in the United States. Cambridge, MA: Health Effects Institute, 2000.

20. Samet JM, Dominici F, Zeger SL, et al. The National Morbidity, Mortality, and Air Pollution Study. Part I: methods and methodologic issues. Cambridge, MA: Health Effects Institute, 2000.

21. EarthInfo, Inc. NCDC Surface Airways. Boulder, CO: EarthInfo, Inc, 1994. (World Wide Web URL: http://www.earthinfo.com/databases/sa.htm).

22. Environmental Protection Agency. Aerometric Information Retrieval System. Washington, DC: Environmental Protection Agency, 1999. (World Wide Web URL: http://www.epa.gov/air/data/info.html).

23. Samet JM, Zeger SL, Berhane K. The association of mortality and particulate air pollution. Part I. Particulate air pollution and daily mortality: replication and validation of selected studies. Cambridge, MA: Health Effects Institute, 1995:1–104.

24. Samet JM, Zeger SL, Kelsall JE, et al. Particulate air pollution and daily mortality: analyses of the effects of weather and multiple air pollutants. The phase IB report of the Particle Epidemiology Evaluation Project. Cambridge, MA: Health Effects Institute, 1997.

25. Samet JM, Dominici F, Curriero FC, et al. Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. N Engl J Med 2000;343:1742–9.

26. Kelsall JE, Samet JM, Zeger SL, et al. Air pollution and mortality in Philadelphia, 1974–1988. Am J Epidemiol 1997;146:750–62.

27. Dominici F, Samet J, Zeger SL. Combining evidence on air pollution and daily mortality from the largest 20 U.S. cities: a hierarchical modeling strategy (with discussion). J R Stat Soc Ser A 2000;163:263–302.

28. Dominici F, Daniels M, Zeger SL, et al. Air pollution and mortality: estimating regional and national dose-response relationships. J Am Stat Assoc 2002;97:100–11.

29. Almon S. The distributed lag between capital appropriations and expenditures. Econometrica 1965;33:178–96.

30. Zanobetti A, Wand M, Schwartz J. Generalized additive distributed lag models. Biostatistics 2000;1:279–92.

31. Daniels MJ, Dominici F, Samet JM, et al. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. Am J Epidemiol 2000;152:397–406.

32. Dominici F, Zeger S, Samet J. A measurement error correction model for time-series studies of air pollution and mortality. Biostatistics 2000;1:157–74.

APPENDIX

Here we outline the approach to decomposing a daily time series X_t into timescale components $\{X_{kt} : \Sigma_{k=1}^{K} X_{kt} = X_t\}$ through the use of the discrete Fourier transform. The discrete Fourier transform is defined as

$$d(\omega_j) = \frac{1}{T}\sum_{t=0}^{T-1} X_t \exp(-i\omega_j t),$$

$$0 \leq j \leq T-1, \quad 0 \leq \omega_j \leq 2\pi,$$

where T is the length of the series X_t and $\omega_j = 2\pi j/T$ is the jth Fourier frequency with j cycles in the length of the data. Note that if $j = 1$, then $\omega_1 = 2\pi/T$ is a Fourier frequency with one cycle in the length of the data and describes the longest-term fluctuations. We note that when $j \geq T/2$, we have $d(\omega_{T-j}) = \overline{d(\omega_j)}$, where $\overline{d(\omega_j)}$ denotes the complex conjugate of $d(\omega_j)$. If T is even and $j = T/2$, then $\omega_{T/2} = \pi$ is a Fourier frequency with a cycle for 2 days and describes the shortest-term fluctuations. Similarly, if T is odd and $j = (T-1)/2$, then

$$\omega_{T-1/2} = \frac{T-1}{T}\pi$$

is the Fourier frequency describing the shortest-term fluctuations.

Let $[0, \omega_1, \ldots, \omega_k, \ldots, \omega_K, \pi]$ be a partition of the interval $[0, \pi]$, and we define $I_k = (\omega_{k-1}, \omega_k] \cup [\omega_{T-k}, \omega_{T-k+1})$. The following holds:

$$X_t = \Sigma_{j=0}^{T-1} d(\omega_j)\exp(i\omega_j t)$$

$$= \Sigma_{k=0}^{K}[\Sigma_{\omega_j \in I_k} d(\omega_j)\exp(i\omega_j t)]$$

$$= \Sigma_{k=0}^{K} X_{kt}.$$

We can decompose the X_t into X_{kt}'s by implementing the following algorithm. For $k = 1, \ldots, K$:

• Taper the data X_t and get $\bar{X}_t$.

• Calculate the discrete Fourier transform of $\bar{X}_t$ and get $d(\omega_j)$.

• Set $d^*(\omega_j) = \begin{cases} d(\omega_j) & \text{for } \omega_j \in I_k \\ 0 & \text{otherwise.} \end{cases}$

• Get X_{kt} by applying the inverse of the discrete Fourier transform to $d^*(\omega_j)$, $j = 1, \ldots, T/2$.

SAS, S-Plus, and R software for decomposing a time series into a desired set of frequency components can be downloaded at http://www.ihapss.jhsph.edu/software/fd/software_fd.htm.

ORIGINAL CONTRIBUTIONS

Airborne Particulate Matter and Mortality: Timescale Effects in Four US Cities

Francesca Dominici[1], Aidan McDermott[1], Scott L. Zeger[1], and Jonathan M. Samet[2]

[1] Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.
[2] Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

While time-series studies have consistently provided evidence for an effect of particulate air pollution on mortality, uncertainty remains as to the extent of the life-shortening implied by those associations. In this paper, the authors estimate the association between air pollution and mortality using different timescales of variation in the air pollution time series to gain further insight into this question. The authors' method is based on a Fourier decomposition of air pollution time series into a set of independent exposure variables, each representing a different timescale. The authors then use this set of variables as predictors in a Poisson regression model to estimate a separate relative rate of mortality for each exposure timescale. The method is applied to a database containing information on daily mortality, particulate air pollution, and weather in four US cities (Pittsburgh, Pennsylvania; Minneapolis, Minnesota; Seattle, Washington; and Chicago, Illinois) from the period 1987–1994. The authors found larger relative rates of mortality associated with particulate air pollution at longer timescale variations (14 days–2 months) than at shorter timescales (1–4 days). These analyses provide additional evidence that associations between particle indexes and mortality do not imply only an advance in the timing of death by a few days for frail individuals.

air pollution; Fourier analysis; hierarchical model; mortality; Poisson distribution; time factors; time series

Abbreviations: CI, confidence interval; PM_{10}, particulate matter with an aerodynamic diameter ≤ 10 μg/m³.

A number of studies over the last decade have shown an association between particle concentrations in outdoor air and daily mortality counts in urban locations (1–3). These associations have been estimated through the use of Poisson regression methods, and the findings have been reported as log relative rates of mortality associated with air pollution levels on recent days. These associations have been widely interpreted as reflecting the effect of air pollution on persons who have heightened susceptibility because of chronic heart or lung diseases (4).

Thus, the increased mortality associated with higher pollution levels may be restricted to very frail people whose life expectancy would have been short even without air pollution. This possibility is termed the "mortality displacement"

Correspondence to Dr. Francesca Dominici, Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205 (e-mail: fdominic@jhsph.edu).

or "harvesting" hypothesis (5). If an effect is evident only at short timescales, pollution-related deaths are advanced by only a few days, and in fact, the days of life lost might arguably be of low quality for the frail individuals at risk of dying. Consequently, the public health relevance of the findings of the daily time-series studies has been questioned (6). The mortality displacement hypothesis received specific discussion in the 1996 Staff Paper on Particulate Matter prepared by the US Environmental Protection Agency because of its policy implications (4). The findings of two long-term prospective cohort studies of air pollution and mortality, the Harvard Six Cities Study (7) and the American Cancer Society's Cancer Prevention Study II (8), were considered to offer critical evidence counter to the mortality displacement hypothesis.

Several investigators have approached the problem of mortality displacement using analytical models for daily time-series data (9–12). If the association between air pollution and mortality does reflect the existence of a pool of frail individuals in the population, episodes of high pollution that lead to increased mortality might reduce the size of this pool, and days subsequent to high-pollution days would then be expected to show a reduced effect of air pollution. Therefore, the occurrence of this phenomenon can be investigated by assessing interaction between prior high-pollution days and the effects of subsequent pollution exposure on mortality counts; under the mortality displacement hypothesis, a negative interaction is predicted (13, 14).

Recently, Kelsall et al. (15) and Schwartz (11) developed related methods for analysis of daily time-series data, both offering approaches to estimating air pollution-mortality associations at varying timescales. More specifically, Kelsall et al.'s methodology gives a continuous smooth estimate of relative risk as a function of timescale (frequency domain log-linear regression). Zeger et al. (10) applied the frequency domain log-linear regression to previously analyzed data for Philadelphia, Pennsylvania, from 1973–1988. Schwartz (11, 12) used a filtering algorithm (16) to separate the time series of daily deaths, air pollution, and weather into long-wavelength components, midscale components, and residual, very short-term components and applied this method to data on Boston, Massachusetts, from 1979–1986 and Chicago, Illinois, from 1988–1993. Note that both of these methods (10, 11) analyze both pollution and mortality on the same timescales, i.e., shorter-term to longer-term. Both sets of analyses found effects on longer timescales.

In this paper, we extend the work by Zeger et al. (10) and Schwartz (11) in the methodological, substantive, and computational arenas. More specifically, we develop a timescale decomposition of a time series based on the discrete Fourier transform; we introduce a two-stage model for combining evidence across locations for estimation of pooled timescale-specific air pollution effects on mortality; and we provide the software for decomposing a time series into a set of desired timescale components. At the first stage of the model, we use Fourier series analyses (17, 18) to decompose the daily time series of the air pollution variable into distinct timescale components. This decomposition leads to a set of orthogonal predictors, each representing a specific timescale of variation in the exposure. We then use this set of predictors in Poisson

regression models to estimate a relative rate of mortality corresponding to each timescale exposure while controlling for other covariates such as temperature. A comparison between our approach and the frequency domain log-linear regression analysis is provided below in the section "Sensitivity analysis and model comparison."

The method is applied to concentrations of particulate matter, based on measurements of particles with an aerodynamic diameter less than or equal to 10 $\mu g/m^3$ (PM_{10}) and daily mortality counts from four US cities—Pittsburgh, Pennsylvania; Minneapolis, Minnesota; Chicago, Illinois; and Seattle, Washington. These were four cities with daily PM_{10} measurements that were among the 90 largest US cities used in the National Morbidity, Mortality, and Air Pollution Study (19, 20). The analyses are restricted to these cities because they are the only US locations with daily air pollution concentrations available in this database for this time interval, while in most other locations, PM_{10} levels were measured only every 6 days as required by the Environmental Protection Agency. Our approach is not suitable for every-sixth-day PM_{10} data, for two reasons: 1) no information is available from the data for estimation of the short-term effects of air pollution on mortality and 2) because of the "aliasing" phenomenon, the effects of air pollution at the longer timescales are distorted. In our context, the aliasing phenomenon occurs when the sampling interval is larger than 1 day, so that variations in the daily time series at the shortest timescales produce an apparent effect at the longer timescales.
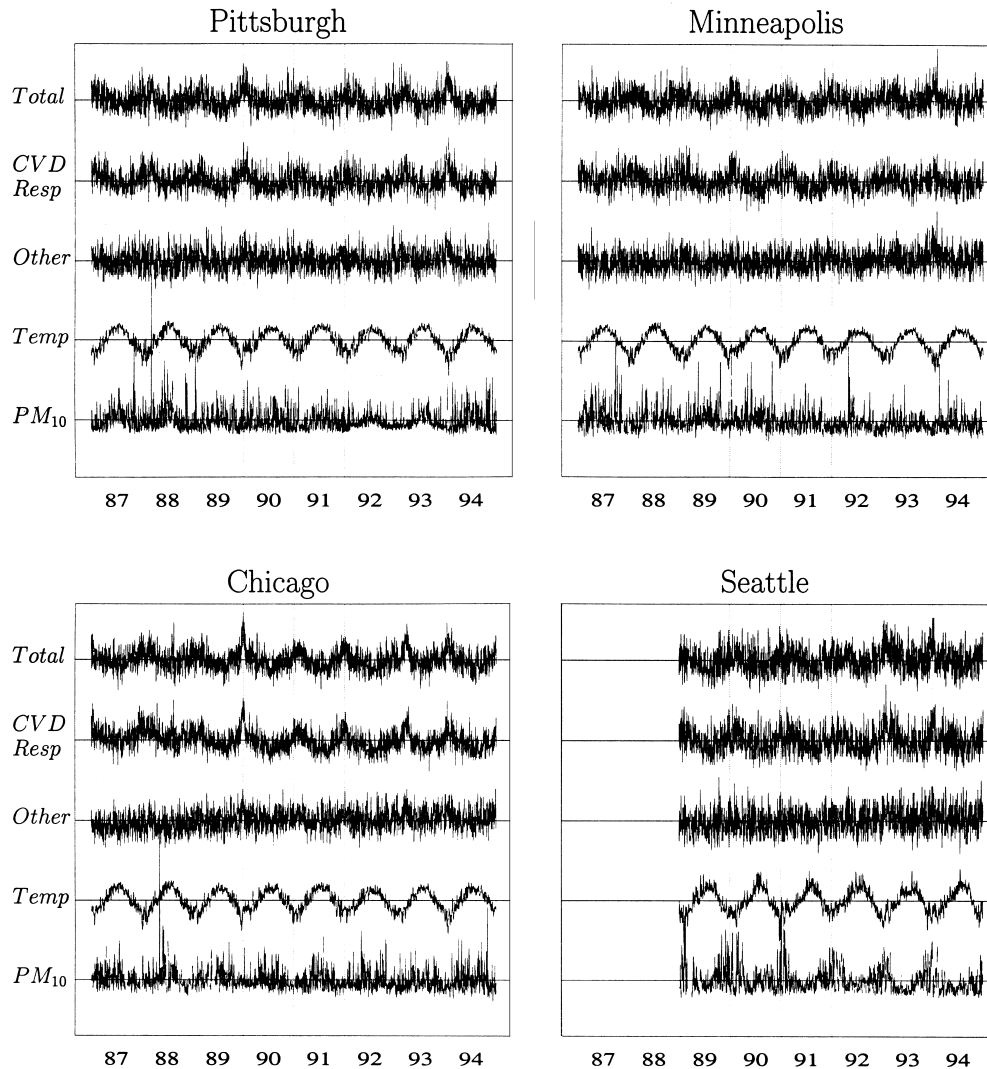
MATERIALS AND METHODS

Data

We used daily time series of mortality, weather, and air pollution data for Pittsburgh, Minneapolis, Chicago, and Seattle for the period 1987–1994 (see figure 1). Daily mortality counts were obtained from the National Center for Health Statistics and were grouped by age (<65, 65–75, and >75 years) and by cause of death according to the *International Classification of Diseases*, Ninth Revision (cardiovascular-respiratory mortality (cardiac conditions, codes 390–448; respiratory conditions, codes 490–496; influenza, code 487; and pneumonia, codes 480–486, 507) and mortality due to other remaining diseases). Accidental deaths were excluded. Hourly temperature and dew point data were available from the National Climatic Data Center, assembled in a compact disk database from EarthInfo, Inc. (21). The air pollution data were obtained from the Aerometric Information Retrieval Service (22) database maintained by the Environmental Protection Agency. For the pollutants measured on an hourly basis, we calculated the 24-hour average. A more detailed description of the database has been published elsewhere (19, 20).

Methods

Below we describe our statistical approach to estimation of the association between air pollution and mortality using different timescales. We let X_t^c be the air pollution time

FIGURE 1. Daily time series of mortality (total, cardiovascular disease (CVD) and respiratory (Resp), and other causes (Other)), temperature (Temp), and levels of particulate matter with an aerodynamic diameter less than 10 µg/m³ (PM_{10}) for Pittsburgh, Pennsylvania, Minneapolis, Minnesota, and Chicago, Illinois, during the period 1987–1994. For Seattle, Washington, data for the period 1989–1994 were used.
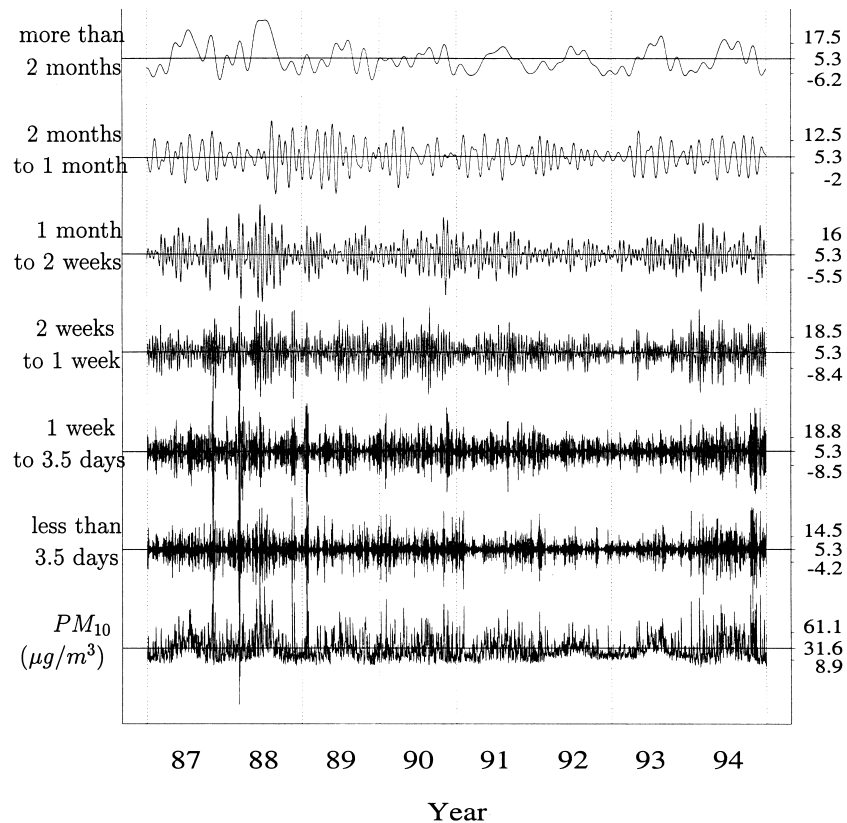
series and Y_t^c be the mortality time series in location c. We first decompose the air pollution series X_t^c into distinct component series X_{kt}^c, one for each distinct timescale k, and then we calculate the association between Y_t^c, without decomposition, and each of the timescale components X_{kt}^c. The decomposition is obtained by applying the discrete Fourier transform to the X_t^c series (17, 18). Specifically, we assume

$$Y_t^c \big| \mu_t^c \sim \text{Poisson}(\mu_t^c); \ \text{var}(Y_t^c) = \varphi^c \mu_t^c$$

and

$$\log \mu_t^c = \Sigma_k X_{kt}^c \beta_k^c + S(\text{time}, \ 7/\text{year}) + \text{confounders}, \quad (1)$$

where φ^c denotes the overdispersion parameter and the β_k^c's, the parameters of interest, denote the log relative rate of daily mortality for each 10-unit increase in the air pollution level in location c on a timescale k. Our modeling approach replaces the term $X_t^c \beta^c$, where X_t^c is the air pollution time series and β^c is the city-specific log relative rate of mortality, with the sum $\Sigma_k X_{kt}^c \beta_k^c$, where $X_t = \Sigma_k X_{kt}$, and the $X_{1t}, \ldots, X_{kt}, \ldots, X_{Kt}$ is a set of orthogonal predictors. This model estimates relative rates of mortality at different timescales and characterizes the timescale variation in the air pollution time series that contributes to the estimate of the overall effect β^c. Here we expect that under a short-term mortality displacement scenario, mortality would be mainly associated with a short-term effect of air pollution.

FIGURE 2. Decomposition into a six-component series of data on particulate matter with an aerodynamic diameter ≤10 µg/m³ (PM₁₀) for Pittsburgh, Pennsylvania, from the period 1987–1994. The bottom row shows the observed PM₁₀ values with a horizontal line plotted at the mean value (31.6 µg/m³) and tick marks at the 10th (8.9 µg/m³) and 90th (61.1 µg/m³) percentiles. Time series 1 to 6 (top to bottom) are the timescale decompositions of the observed data series. A horizontal line is plotted at the mean value, which in all cases is one sixth of the original mean (5.3 µg/m³), and tick marks show the 10th and 90th percentiles (see vertical axis at right). The decomposition is obtained by applying the function *decompose*( ) with breaks equal to 1, 48, 96, 208, 416, 834, and 2,922 days. The internal breaks (48, 96, 208, 416, and 834) were obtained by dividing the length of the data (e.g., *T* = 2,922 days) by timescales of 60, 30, 14, 7, and 3.5 days. One should always select the minimum and maximum breaks to be 1 and *T* in order to capture the lowest and highest Fourier frequencies.

To protect the pollution relative rates from confounding by longer-term trends and seasonality, we also remove the variation in the time series at timescales approximately longer than 2 months by including a smooth function of time with 7 degrees of freedom per year. A sensitivity analysis with respect to selection of the number of degrees of freedom in the smooth function of time is discussed below. Smooth functions of temperature and dew point temperature are used to control for potential confounding by temperature and humidity. The rationale for and details on the selected smooth functions are provided by Samet et al. (23–25), Kelsall et al. (26), and Dominici et al. (27).
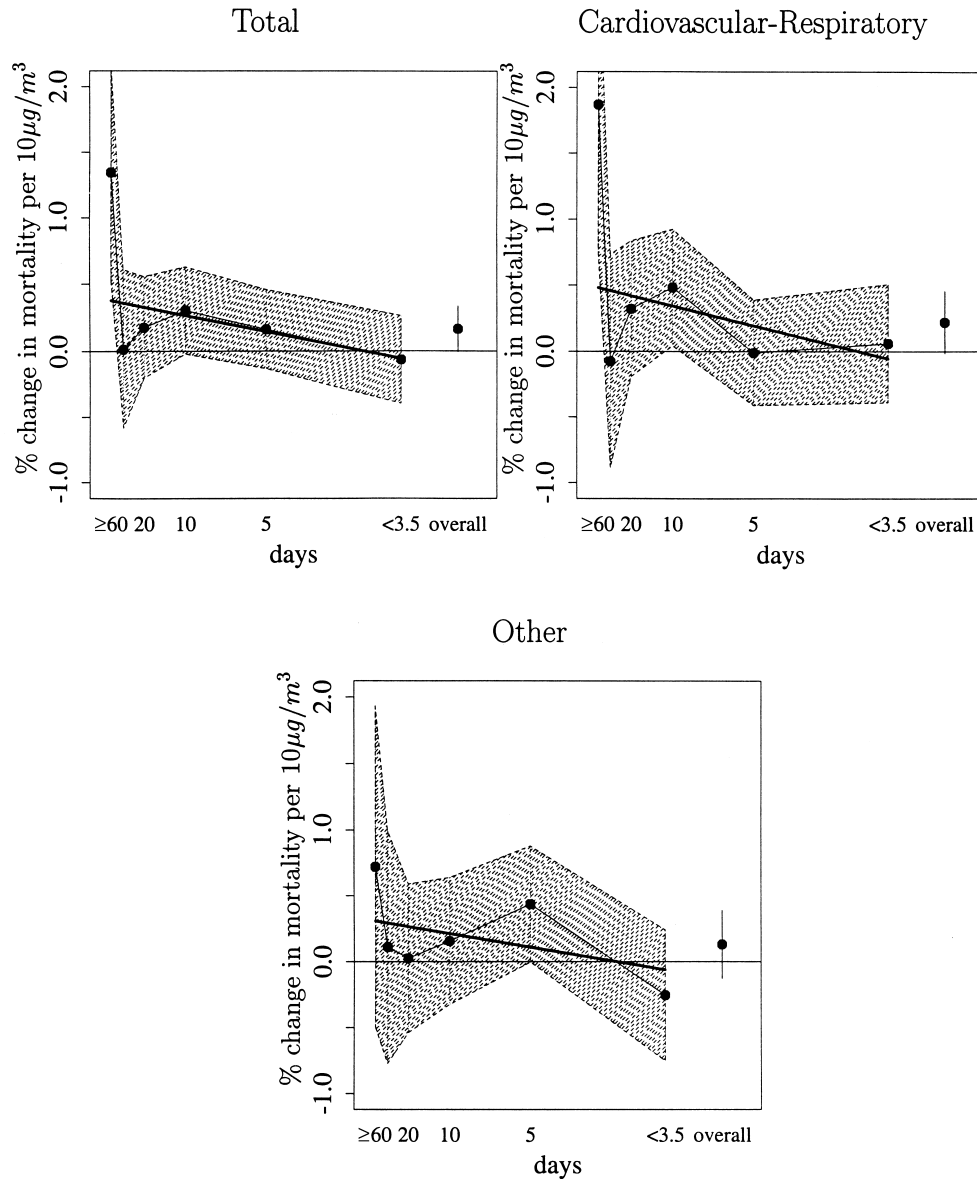
Figure 2 illustrates the decomposition of X_t into six timescales. From the top of the panel to the bottom are displayed time series ranging from series that comprise only the more smooth fluctuations (low frequency components) to time series that comprise only the less smooth variations (high frequency components). The actual value of X_t (the last time series at the bottom) on day t is obtained by summing the values of the six component series on each day. Details on the Fourier series decomposition and the URL address for downloading the software for its implementation are provided in the Appendix. Using the decomposed time series, we can estimate timescale coefficients ($\hat{\beta}_k^c$) denoting the relative change in mortality per 10-µg/m³ increase in the corresponding timescale components of PM₁₀ (X_{kt}^c). We estimate a vector of regression coefficients $\hat{\beta}^c = [\hat{\beta}_1^c, \ldots, \hat{\beta}_K^c]$ and their covariance matrix V^c. Although one might want to allow a latency time for the effect of pollution, in equation 1 we regress Y_t^c on X_{kt}^c rather than use lagged pollution series $X_{kt-l_k}^c$ for some lag $l_k > 0$. We investigate whether a lagged predictor is needed in the sensitivity analysis.

RESULTS

Estimation

To estimate the timescale effects $\hat{\beta}^c$ and its covariance matrix V^c within each location c, we fit the Poisson regression model (equation 1) with smooth functions of time, temperature, and dew point modeled as natural cubic splines and using the function glm( ) in S-Plus (MathSoft, Inc., Cambridge, Massachusetts). To combine the coefficients

FIGURE 3. Pooled estimates of log relative rates of mortality at different timescales for total mortality (top left), cardiovascular-respiratory mortality (top right), and mortality due to other causes (bottom) according to levels of particulate matter with an aerodynamic diameter ≤ 10 µg/m³ in four US cities, 1987–1994. The estimates are plotted at ≥ 60, 40 (not shown on x-axis), 20, 10, 5, and <3.5 days, denoting approximate midpoints of the intervals ≥ 60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days, respectively. The shaded regions represent ±2 standard errors of the estimates. Bars, 95% confidence interval.

across cities, we use a fixed-effect model with weights $W^c = (V^c)^{-1}$ and an estimator of the form

$$\bar{\beta} = \left(\Sigma_c W^c \right)^{-1} \sum_c W^c \hat{\beta}^c$$

with variance

$$V(\bar{\beta}) = \left(\Sigma_c W^c \right)^{-1}.$$

An alternative approach would be to use as weights $W^c = (D + V^c)^{-1}$, where D is a diagonal between-city covariance matrix with diagonal element τ^2. Because of the limited number of cities in the present analysis, we cannot estimate τ^2 reliably and have assumed $\tau = 0$. A sensitivity analysis of our results with respect to different values of τ^2 obtained from hierarchical analyses of data from 20 cities (27) and 88 cities (28) is discussed below.

We estimated city-specific and pooled log relative rates of mortality for the following six timescale variations of PM_{10}: ≥ 60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days. Figure 3 shows the pooled estimates of the log relative rates of mortality at different timescales for total

TABLE 1. Timescale and city-specific estimates of log relative rates of mortality (total, cardiovascular and respiratory, and other causes) associated with a 10-unit increase in particulate matter with an aerodynamic diameter less than or equal to 10 μg/m³ in four US cities, 1987–1994

Type of mortality and timescale* (days)	Pittsburgh, Pennsylvania		Minneapolis, Minnesota		Chicago, Illinois		Seattle, Washington		Pooled data	
	Log RR†	95% CI†	Log RR	95% CI	Log RR	95% CI	Log RR	95% CI	Log RR	95% CI
Total mortality										
≥60	1.34	0.18, 2.86	0.51	−1.90, 2.93	1.57	0.37, 2.77	1.37	−0.98, 3.72	1.35	0.52, 2.17
40	0.95	−0.23, 2.13	−1.11	−2.97, 0.75	−0.14	−0.97, 0.69	−0.33	−1.98, 1.32	0.01	−0.58, 0.61
20	−0.57	−1.30, 0.15	1.21	−0.10, 2.51	0.36	−0.15, 0.87	0.38	−0.92, 1.69	0.18	−0.20, 0.56
10	0.19	−0.45, 0.83	0.06	−1.03, 1.15	0.42	−0.01, 0.84	0.11	−1.19, 1.40	0.30	−0.02, 0.63
5	0.27	−0.27, 0.81	0.92	−0.12, 1.95	−0.01	−0.40, 0.37	0.43	−1.15, 2.00	0.16	−0.13, 0.46
<3.5	0.09	−0.57, 0.75	−0.67	−1.86, 0.52	−0.06	−0.47, 0.36	0.02	−2.30, 2.33	−0.07	−0.40, 0.27
Overall	0.15	−0.19, 0.49	0.24	−0.37, 0.84	0.16	−0.07, 0.38	0.26	−0.45, 0.98	0.17	−0.01, 0.34
Cardiovascular and respiratory mortality										
≥60	1.19	−0.86, 3.24	0.99	−2.36, 4.35	2.54	0.90, 4.17	1.86	−1.44, 5.15	1.87	0.75, 2.99
40	1.00	−0.59, 2.59	−1.80	−4.38, 0.78	0.35	−0.78, 1.48	−2.67	−4.95, −0.39	−0.08	−0.89, 0.74
20	−0.33	−1.31, 0.64	1.00	−0.80, 2.79	0.45	−0.24, 1.14	1.04	−0.76, 2.85	0.32	−0.19, 0.84
10	0.44	−0.42, 1.31	−0.12	−1.62, 1.38	0.63	0.05, 1.20	0.24	−1.56, 2.03	0.48	0.04, 0.92
5	0.22	−0.50, 0.95	0.51	−0.92, 1.94	−0.24	−0.77, 0.28	0.84	−1.34, 3.01	−0.01	−0.41, 0.39
<3.5	0.23	−0.65, 1.11	−0.48	−2.13, 1.16	0.05	−0.51, 0.61	0.22	−2.99, 3.44	0.06	−0.39, 0.51
Overall	0.25	−0.20, 0.70	0.07	−0.77, 0.90	0.23	−0.08, 0.53	0.22	−0.76, 1.21	0.22	−0.02, 0.46
Mortality due to other causes										
≥60	1.45	−0.85, 3.74	−0.83	−4.36, 2.69	0.75	−1.02, 2.52	0.56	−2.83, 3.96	0.72	−0.50, 1.94
40	0.85	−0.92, 2.61	0.02	−2.70, 2.73	−0.74	−1.98, 0.50	2.04	−0.38, 4.47	0.11	−0.77, 0.99
20	−0.93	−2.01, 0.16	1.26	−0.65, 3.17	0.31	−0.44, 1.07	−0.23	−2.13, 1.68	0.03	−0.54, 0.59
10	−0.10	−1.06, 0.86	0.33	−1.26, 1.93	0.22	−0.41, 0.84	0.30	−1.59, 2.19	0.16	−0.33, 0.64
5	0.40	−0.42, 1.21	1.39	−0.13, 2.90	0.32	−0.25, 0.90	0.28	−2.02, 2.58	0.44	0.00, 0.88
<3.5	−0.14	−1.14, 0.86	−1.03	−2.77, 0.72	−0.20	−0.81, 0.42	−0.45	−3.82, 2.91	−0.26	−0.75, 0.24
Overall	0.02	−0.49, 0.53	0.38	−0.50, 1.27	0.11	−0.23, 0.45	0.38	−0.66, 1.43	0.13	−0.13, 0.39

* The numbers given for the timescales denote the midpoints of the following intervals: ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days.
† RR, relative rate; CI, confidence interval.

mortality, cardiovascular and respiratory mortality, and mortality due to other causes. At the far right are the plotted estimates of the log relative rate of mortality obtained using the nondecomposed time series X_t. For all causes and for cause-specific mortality, we found that estimates of the association between air pollution and mortality obtained using the smoother variations in the time series (10 days to 1 month) are somewhat larger than those obtained using the less smooth variations (1–3 days). The largest effects occurred at timescales greater than 2 months for total mortality (1.35 percent per 10 μg/m³; 95 percent confidence interval (CI): 0.52, 2.17), cardiovascular and respiratory mortality (1.87 percent per 10 mg/m³; 95 percent CI: 0.75, 2.99), and other-cause mortality (0.72 percent per 10 μg/m³; 95 percent CI: –0.55, 1.95). To test the hypothesis that estimated effects at the longer timescales are larger than those at the shortest timescales, we linearly regress the

pooled $\hat{\beta}_k$ on the timescales and calculated weighted least squares estimates. The solid lines in figure 3 represent the fitted linear regressions. In all cases, the estimated slopes are negative, with t statistics close to the significance level.

Table 1 shows the estimates of the timescale effect for each city. At the city-specific level, we found a similar pattern, with larger effects at the longer timescales and smaller effects at the shorter timescales, with few exceptions. Consistent with previous studies (19, 20), the log relative rates, obtained by averaging across cities and across timescales, are highest for cardiovascular and respiratory mortality (0.22 percent per 10 μg/m³; 95 percent CI: –0.02, 0.46) and lowest for other-cause mortality (0.13 percent per 10 μg/m³; 95 percent CI: –0.13, 0.39). These findings suggest that the estimates of association between air pollution and mortality at the greater timescales contribute substantially to the estimation of the overall log relative rate β^c.

TABLE 2.   Analysis of the sensitivity of estimated log relative rates of mortality due to air pollution to the choice of lag for exposure series at timescales shorter than 1 month in four US cities, 1987–1994*

City	14–29 days	7–13 days	3.5–6 days	<3.5 days
Pittsburgh, Pennsylvania	1	2	0	6
Minneapolis, Minnesota	0	5	2	3
Chicago, Illinois	0	0	6	5
Seattle, Washington	0	3	6	4

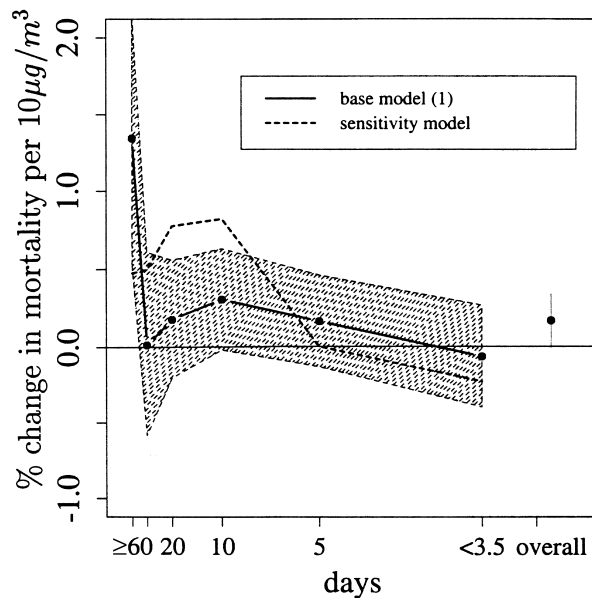* The table summarizes, for each city and each timescale, the lag associated with the largest *t* statistic.

Sensitivity analysis and model comparison

Below we investigate the sensitivity of our results with respect to: 1) lag choice in the air pollution time series; 2) adjustment for long-term trends and seasonality; and 3) the degree of heterogeneity of the true relative rates among cities. We also apply our timescale approach to the Philadelphia database for 1973–1988 that was previously analyzed by Zeger et al. (10), to compare methods used here with the frequency domain log-linear regression estimates previously published.

We first test the sensitivity of the log relative rate estimates to the choice of lag for component exposure series at timescales shorter than 1 month. We assume that the lag l_k is 0 for timescales greater than 1 month, since lags of 4 days will have little effect on the results for large timescales. We fit several different lags for each component exposure series X_{kt-l_k} and choose the best lag l_k (the one with the largest *t* statistic) rather than assume that l_k is 0. The optimal lags were obtained by including all timescale components in the models; they are summarized in table 2. Results for total mortality with an optimal lag compared with the original model with a zero lag are shown in the upper half of figure 4. Although the estimates differ at particular timescales, the overall shape of the curves remains similar and remains inconsistent with the short-term mortality displacement hypothesis.

Our model controls for long-term trends in mortality by including a natural cubic spline of time with 7 degrees of freedom per year. To assess the sensitivity of the results to the choice of smoothing parameter, we repeat the analysis using 3.5 and 14 degrees of freedom per year. The lower half of figure 4 shows the resulting plots. As expected, when less flexible curves are used (fewer degrees of freedom), there is less control for trend and seasonality, and therefore the effects at the longer timescales tend to be higher.

Our strategy for investigating the impact of the assumption of homogeneity ($\tau = 0$) of the pollution effects on our results is based on inspecting the pooled timescale estimates for total mortality under four alternative values for τ. These were extracted from Bayesian hierarchical analyses of 20 cities (27) and the 88 largest cities in the United States (28). The posterior mean values of τ and the corresponding prior distributions are summarized in table 3.

FIGURE 4.   Top: Sensitivity analysis showing stepwise pooled estimates for the total mortality and PM$_{10}$ (particulate matter with an aerodynamic diameter ≤10 µg/m^3) data series (four US cities, 1987–1994) and model 1 (see equation 1 in text). Bottom: Sensitivity analysis showing pooled estimates for the total mortality and PM$_{10}$ data series using different numbers of degrees of freedom (df) for time but the same smooth function for temperature and humidity. The estimates are plotted at ≥60, 40 (not shown on *x*-axis), 20, 10, 5, and <3.5 days, denoting approximate midpoints of the intervals ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days, respectively. Bars, 95% confidence interval.

Results are shown in table 4. The results were all obtained by using smooth functions (natural cubic splines) of time with 7 degrees of freedom per year and smooth functions (natural cubic splines) of temperature and dew point as

TABLE 3. Sensitivity analysis of pooled estimates of the log relative rate of mortality due to air pollution with respect to the amount of heterogeneity (τ)*

$E[\tau|\text{data}]$	Prior distribution	Published reference
0.15	$\tau^2 \sim N\dagger(0,\ 305)\,I_{\tau^2 > 0}$	20-city analysis (discussion and rejoinder of Dominici et al. (27))
0.38	$\tau^2 \sim N(0,\ 1)\,I_{\tau^2 > 0}$	88-city analysis (Dominici et al. (28))
0.49	$\tau^2 \sim IG\ddagger(3,\ 1)$	88-city analysis (Dominici et al. (28))
0.76	$\tau^2 \sim IG(3,\ 6)$	20-city analysis (Dominici et al. (27))

* The table summarizes posterior mean values of τ used to calculate the pooled effect.
† N denotes the normal distribution.
‡ IG denotes the inverse gamma distribution.

confounders, as in model 1 (equation 1). Although the confidence intervals widen considerably, the pooled estimates are not very sensitive to the different values of τ. Even in the presence of substantial heterogeneity of the relative rates of mortality across cities, we still found that the pooled estimates at the longer timescales are larger than the pooled estimates at the shorter timescales, but with larger standard errors.

We now apply timescale and frequency domain log-linear regression analyses (10, 15) to the Philadelphia data set, and we estimate relative rates of mortality for exposure to air pollution at different timescales by using the Poisson regression model defined in equation 1. Figure 5 shows the frequency domain estimate of the mortality relative rate associated with air pollution as a function of Fourier frequency. Similar to the timescale result, the horizontal axes denote the Fourier frequencies (lower *x*-axis) and the timescale in days (upper *x*-axis) at which the association is measured. The solid curve and dotted curves denote the estimated relative rates ±2 estimated standard errors at each frequency. The timescale estimates (points connected by line segments) are plotted on top of the frequency domain results (continuous curve). Timescale estimates and frequency domain results are similar, and consistently with our results for the four cities, relative rate estimates at longer timescales are larger than relative rate estimates at short timescales.

DISCUSSION

This paper provides additional evidence that the association between particle indexes and mortality is greater at longer timescales (10 days to 2 months) than at timescales of a few days. This suggests that the association of air pollution with daily mortality counts does not reflect short-term mortality displacement alone. More specifically, our results are inconsistent with the "harvesting only" hypothesis, which contends that the air pollution-mortality association is caused entirely by frail persons' dying a few days earlier than they would have absent pollution. Under that hypothesis, we would anticipate little or no association at longer timescales. In fact, we observe the strongest associations there.

The larger relative rates at longer timescales may partly reflect a greater biologic impact on chronic exposures than on acute exposures. In fact, estimated relative risks from the Harvard Six Cities (7) and American Cancer Society (8) cohort studies, which address chronic exposures, are larger than estimates from times-series models (28), which are constrained to estimate the effects of shorter-timescale exposures.
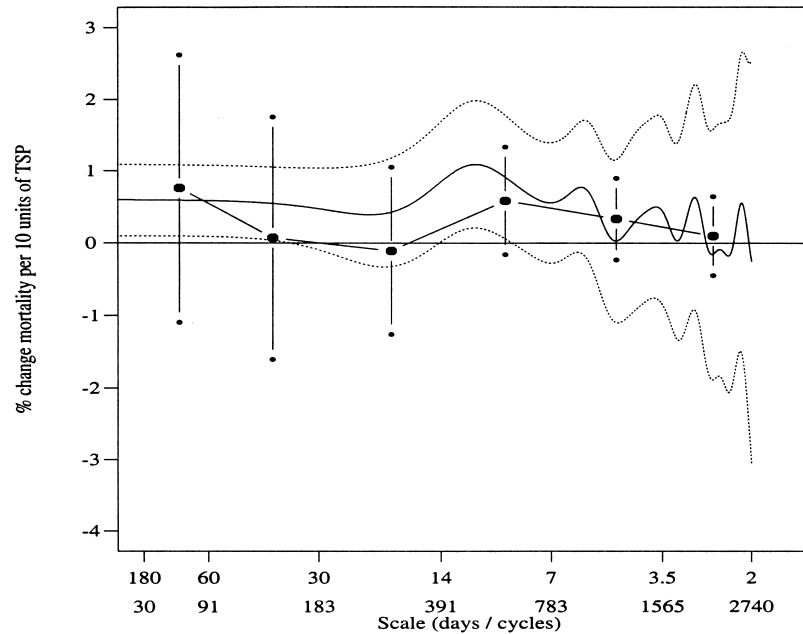
The estimated relative rate of total mortality for the longest timescale (2 months) was 1.35 percent per 10-unit increase in PM_{10} for the four cities considered. While 1.35 percent is approximately 8 times larger than the overall pooled esti-

TABLE 4. Pooled estimates of the log relative rate of mortality due to air pollution under different values of the heterogeneity parameter τ

Timescale* (days)	$\tau = 0$ Log RR†	95% CI†	$\tau = 0.15$ Log RR	95% CI	$\tau = 0.38$ Log RR	95% CI	$\tau = 0.49$ Log RR	95% CI	$\tau = 0.76$ Log RR	95% CI
≥60	1.35	0.52, 2.17	1.34	0.50, 2.18	1.32	0.39, 2.25	1.31	0.32, 2.30	1.29	0.12, 2.45
40	0.01	−0.58, 0.61	0.01	−0.61, 0.63	0.00	−0.74, 0.73	−0.02	−0.83, 0.79	−0.05	−1.06, 0.96
20	0.18	−0.20, 0.56	0.18	−0.24, 0.60	0.21	−0.36, 0.79	0.24	−0.43, 0.90	0.28	−0.61, 1.18
10	0.30	−0.02, 0.63	0.28	−0.09, 0.66	0.24	−0.31, 0.79	0.23	−0.41, 0.87	0.21	−0.66, 1.09
5	0.16	−0.13, 0.46	0.20	−0.15, 0.55	0.29	−0.25, 0.83	0.32	−0.32, 0.95	0.36	−0.52, 1.23
<3.5	−0.07	−0.40, 0.34	−0.07	−0.46, 0.32	−0.11	−0.69, 0.47	−0.12	−0.80, 0.56	−0.14	−1.07, 0.78
Overall	0.17	−0.01, 0.34	0.18	−0.07, 0.43	0.19	−0.26, 0.64	0.19	−0.35, 0.74	0.20	−0.60, 1.00

* The numbers given for the timescale denote the midpoints of the following intervals: ≥60 days, 30–59 days, 14–29 days, 7–13 days, 3.5–6 days, and <3.5 days.
† RR, relative rate; CI, confidence interval.

FIGURE 5. Change in mortality according to level of total suspended particulates (TSP) (particulate matter with an aerodynamic diameter ≤10 µg/m³) in Philadelphia, Pennsylvania, 1973–1988. The figure depicts a comparison between frequency domain estimates (continuous curve) and time-scale estimates (points connected by line segments), showing the log relative rates of total mortality by frequency and frequency grouping. The dotted lines show ±2 standard errors for the frequency domain estimates, and the bars represent ±2 standard errors for the timescale estimates.

mate of 0.17 percent, it is still an order of magnitude smaller than the estimated relative risks from the cohort studies (7, 8). Thus, the time-series relative rates, even when restricted to longer-term exposures, are much smaller than those from the major cohort studies. This difference might indicate that the most harmful exposures occur over much larger timescales than can be studied with time-series methods. However, relative rate estimates at the longer timescales should be interpreted with caution because of the confounding effects of seasonality and trend.

Our results are consistent with findings from previous reports for Philadelphia (10), Boston (11), and Chicago (12) that have used harvesting-resistant estimators. These methods are based on a conceptually straightforward stratification of the air pollution time series into different frequency bands, allowing assessment of associations on timescales with differing implications.

Our approach and the approaches proposed by Zeger et al. (10) and Schwartz (11, 12) address related but different questions. Zeger et al. (10) and Schwartz (11, 12) decompose both the air pollution time series and the mortality time series into different timescales of variation (X_{kt} and Y_{kt}) and then aim to identify the timescale component that leads to the strongest association between time-averaged air pollution and time-averaged mortality. The timescale analysis proposed in this paper decomposes only the air pollution time series into different timescales (X_{kt}) and then characterizes the timescale variation of the effect of exposure on daily mortality. For linear models, these two approaches will provide the same

results. In Poisson regression, with small effects such as those that occur with air pollution variables, the differences between results from the two approaches will probably be small. Our approach, however, is applicable over the range of Poisson or other generalized linear model applications.

The timescale decomposition shown in figure 2 could have been performed using wavelet methods. Wavelets are a natural extension of Fourier analysis; however, in wavelet analysis, the window or "scale" with which we look at the information stream is selected automatically. In our context, this automatic selection of the timescales is not particularly desirable. One of the advantages of using wavelets is that functions with discontinuities and functions with sharp spikes can be represented using substantially fewer wavelet basis elements than sine-cosine basis elements. Because a common characteristic of time series of mortality, air pollution, and weather data is their periodicity without large discontinuities, Fourier analysis is adequate for our purpose.

The mortality displacement problem that motivated the development of this method is not unique to air pollution; it has also been discussed in relation to heat waves and influenza. The statistical approach proposed in this paper is suitable for these or other epidemiologic analyses with the focus of differentiating short-term effects from long-term effects of a time-varying exposure on a health outcome. The set of timescales selected should match hypotheses concerning relations between exposure time and response. We also provide an alternative strategy with which to control for

temporal confounding, since it is likely that confounding may vary with the timescale.

The timescale estimates from model 1 lead to specific patterns for the coefficients of a distributed lag model (9, 29, 30). A large effect at timescale k corresponds to an increased number of deaths for $k/2$ days after an air pollution episode, followed by a rebound below the baseline level for another $k/2$ days, owing to the depletion of the pool of susceptible people.

Unlike the distributed lag model, our approach is symmetric in time; that is, we use a symmetric time window $(t - l_k, t + l_k)$ to estimate X_{kt}. The temporal symmetry of our approach does not complicate our inferences, for two reasons. First, and most importantly, it is not plausible that mortality causes air pollution; it is only reasonable to consider the possibility that air pollution causes mortality. Second, we use a symmetric time window simply to better estimate the smooth variations of air pollution X_{kt}.

Other key methodological issues in time-series studies of air pollution and mortality are the nonlinearity in the dose-response curves, the effect of copollutants, and the effect of measurement error. These issues are discussed elsewhere and remain a topic of investigation (25, 28, 31, 32). In the context of mismeasurement of exposure, it is expected that the relative rate of mortality corresponding to the short time scales might be more attenuated by the measurement error than the relative rate of mortality corresponding to longer timescales. This is because more of the short timescale signal is actually error, whereas the longer timescale measure has effectively been smoothed so that measurement error is less of a contributor and hence less a source of bias. However, measurement error will not reverse the sign of an estimated coefficient or reverse the shape of the curve in figure 3. Therefore, even in the presence of measurement error, our results still do not support the "harvesting hypothesis" that the association between particle concentrations and mortality is entirely due to mortality among very frail persons who lose a few days of life.

ACKNOWLEDGMENTS

REFERENCES

1. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. Health effects of outdoor air pollution. [Part 1]. Am J Respir Crit Care Med 1996;153:3–50.
2. Committee of the Environmental and Occupational Health Assembly of the American Thoracic Society. Health effects of outdoor air pollution. Part 2. Am J Respir Crit Care Med 1996;153:477–98.
3. Pope CA III, Dockery DW, Schwartz J. Review of epidemiological evidence of health effects of particulate air pollution. Inhal Toxicol 1995;7:1–18.
4. Environmental Protection Agency, Office of Air Quality Planning and Standards. Review of the National Ambient Air Quality Standards for Particulate Matter: policy assessment of scientific and technical information. OAQPS Staff Paper. Research Triangle Park, NC: Environmental Protection Agency, 1996. (Publication no. EPA-452\R-96-013).
5. Schimmel H, Murawski TJ. Proceedings: the relation of air pollution to mortality. J Occup Med 1976;18:316–33.
6. Lipfert FW, Wyzga RE. Air pollution and mortality: issues and uncertainties. J Air Waste Manage Assoc 1995;45:949–66.
7. Dockery DW, Pope CA III, Xu X, et al. An association between air pollution and mortality in six U.S. cities. N Engl J Med 1993;329:1753–9.
8. Pope CA III, Thun MJ, Namboordiri MM, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. Am J Respir Crit Care Med 1995;151:669–74.
9. Zanobetti A, Schwartz J, Samoli E, et al. The temporal pattern of mortality responses to air pollution: a multicity assessment of mortality displacement. Epidemiology 2002;13:87–93.
10. Zeger SL, Dominici F, Samet J. Harvesting-resistant estimates of air pollution effects on mortality. Epidemiology 1999;10:171–5.
11. Schwartz J. Harvesting and long term exposure effects in the relationship between air pollution and mortality. Am J Epidemiol 2000;151:440–8.
12. Schwartz J. Is there harvesting in the association of airborne particles with daily deaths and hospital admissions? Epidemiology 2001;12:55–61.
13. Spix C, Heinrich J, Dockery D, et al. Air pollution and daily mortality in Erfurt, East Germany, 1980–1989. Environ Health Perspect 1993;101:518–26.
14. Smith RL, Davis JM, Speckman P. Human health effects of environmental pollution in the atmosphere. In: Barnett V, Stein A, Turkman F, eds. Statistics in the environment 4: statistical aspects of health and the environment. Chichester, United Kingdom: John Wiley and Sons Ltd, 1999:91–115.
15. Kelsall J, Zeger S, Samet J. Frequency domain log-linear models: air pollution and mortality. Appl Stat 1999;48:331–44.
16. Cleveland WS, Develin SJ. Robust locally-weighted regression and smoothing scatterplots. J Am Stat Assoc 1988;74:829–36.
17. Bloomfield P. Fourier analysis of time series: an introduction. New York, NY: John Wiley and Sons, Inc, 1976.
18. Priestley MB. Spectral analysis and time series. New York, NY: Academic Press, Inc, 1981.
19. Samet JM, Zeger SL, Dominici F, et al. The National Morbidity, Mortality, and Air Pollution Study. Part II: morbidity and

mortality from air pollution in the United States. Cambridge, MA: Health Effects Institute, 2000.

20. Samet JM, Dominici F, Zeger SL, et al. The National Morbidity, Mortality, and Air Pollution Study. Part I: methods and methodologic issues. Cambridge, MA: Health Effects Institute, 2000.

21. EarthInfo, Inc. NCDC Surface Airways. Boulder, CO: EarthInfo, Inc, 1994. (World Wide Web URL: http://www.earth-info.com/databases/sa.htm).

22. Environmental Protection Agency. Aerometric Information Retrieval System. Washington, DC: Environmental Protection Agency, 1999. (World Wide Web URL: http://www.epa.gov/air/data/info.html).

23. Samet JM, Zeger SL, Berhane K. The association of mortality and particulate air pollution. Part I. Particulate air pollution and daily mortality: replication and validation of selected studies. Cambridge, MA: Health Effects Institute, 1995:1–104.

24. Samet JM, Zeger SL, Kelsall JE, et al. Particulate air pollution and daily mortality: analyses of the effects of weather and multiple air pollutants. The phase IB report of the Particle Epidemiology Evaluation Project. Cambridge, MA: Health Effects Institute, 1997.

25. Samet JM, Dominici F, Curriero FC, et al. Fine particulate air pollution and mortality in 20 U.S. cities, 1987–1994. N Engl J Med 2000;343:1742–9.

26. Kelsall JE, Samet JM, Zeger SL, et al. Air pollution and mortality in Philadelphia, 1974–1988. Am J Epidemiol 1997;146: 750–62.

27. Dominici F, Samet J, Zeger SL. Combining evidence on air pollution and daily mortality from the largest 20 U.S. cities: a hierarchical modeling strategy (with discussion). J R Stat Soc Ser A 2000;163:263–302.

28. Dominici F, Daniels M, Zeger SL, et al. Air pollution and mortality: estimating regional and national dose-response relationships. J Am Stat Assoc 2002;97:100–11.

29. Almon S. The distributed lag between capital appropriations and expenditures. Econometrica 1965;33:178–96.

30. Zanobetti A, Wand M, Schwartz J. Generalized additive distributed lag models. Biostatistics 2000;1:279–92.

31. Daniels MJ, Dominici F, Samet JM, et al. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. Am J Epidemiol 2000;152:397–406.

32. Dominici F, Zeger S, Samet J. A measurement error correction model for time-series studies of air pollution and mortality. Biostatistics 2000;1:157–74.

APPENDIX

Here we outline the approach to decomposing a daily time series X_t into timescale components $\{X_{kt} : \Sigma_{k=1}^{K} X_{kt} = X_t\}$ through the use of the discrete Fourier transform. The discrete Fourier transform is defined as

$$d(\omega_j) = \frac{1}{T}\sum_{t=0}^{T-1} X_t \exp(-i\omega_j t),$$

$$0 \le j \le T-1, \quad 0 \le \omega_j \le 2\pi,$$

where T is the length of the series X_t and $\omega_j = 2\pi j/T$ is the jth Fourier frequency with j cycles in the length of the data. Note that if $j = 1$, then $\omega_1 = 2\pi/T$ is a Fourier frequency with one cycle in the length of the data and describes the longest-term fluctuations. We note that when $j \ge T/2$, we have $d(\omega_{T-j}) = \overline{d(\omega_j)}$, where $\overline{d(\omega_j)}$ denotes the complex conjugate of $d(\omega_j)$. If T is even and $j = T/2$, then $\omega_{T/2} = \pi$ is a Fourier frequency with a cycle for 2 days and describes the shortest-term fluctuations. Similarly, if T is odd and $j = (T-1)/2$, then

$$\omega_{T-1/2} = \frac{T-1}{T}\pi$$

is the Fourier frequency describing the shortest-term fluctuations.

Let $[0, \omega_1, \ldots, \omega_k, \ldots, \omega_K, \pi]$ be a partition of the interval $[0, \pi]$, and we define $I_k = (\omega_{k-1}, \omega_k] \cup [\omega_{T-k}, \omega_{T-k+1})$. The following holds:

$$X_t = \Sigma_{j=0}^{T-1} d(\omega_j)\exp(i\omega_j t)$$

$$= \Sigma_{k=0}^{K}[\Sigma_{\omega_j \in I_k} d(\omega_j)\exp(i\omega_j t)]$$

$$= \Sigma_{k=0}^{K} X_{kt}.$$

We can decompose the X_t into X_{kt}'s by implementing the following algorithm. For $k = 1, \ldots, K$:

• Taper the data X_t and get $\overline{X}_t$.

• Calculate the discrete Fourier transform of $\overline{X}_t$ and get $d(\omega_j)$.

$$\bullet \text{ Set } d^*(\omega_j) = \begin{cases} d(\omega_j) & \text{for } \omega_j \in I_k \\ 0 & \text{otherwise.} \end{cases}$$

• Get X_{kt} by applying the inverse of the discrete Fourier transform to $d^*(\omega_j)$, $j = 1, \ldots, T/2$.

SAS, S-Plus, and R software for decomposing a time series into a desired set of frequency components can be downloaded at http://www.ihapss.jhsph.edu/software/fd/software_fd.htm.